Causal inference, time and observation plans in the social sciences

Blossfeld, Hans-Peter; Rohwer, Götz

Sonderforschungsbereich 186
der Universität Bremen

Statuspassagen und Risikolagen
im Lebensverlauf

Causal Inference, Time and Observation Plans in the Social Sciences

von

Hans-Peter Blossfeld
und
Götz Rohwer

Arbeitspapier Nr. 36

Bremen 1996

Preface

The B6 project within the Special Collaborative Program on "Status Passages and Risks in the Life Course" uses advanced quantitative longitudinal methodolgies. A great deal of work in this project is directed toward the development of new techniques and methods.

This paper demonstrates that the opportunity for assessing causal inferences varies strongly with the type of observation available to the social scientist. The data structure (which can be cross-sectional, panel or event oriented) determines the extent to which the researcher is forced to make untested assumptions in the process of establishing relevant empirical evidence that can serve as a link in a chain of reasoning about causal mechanisms.

The authors of this paper stress that the collection of event history data offer a comparatively superior approach for uncovering causal relationships or mapping out systems of causal relations. This is because event history models relate the change in future outcomes to conditions in the past at each point in time.

Prof. Dr. Ansgar Weymann
Chai, Special Collaborative Programme No. 186

Table of Contents

1. Introduction

The investigation of causal relationships is an important but difficult scientific endeavor. The opportunity for assessing causal inferences varies strongly with the type of observation available to the social scientist. This is because the data structure determines the extent to which the researcher is forced to make untested assumptions when he or she is trying to establish relevant empirical evidence that can serve as a link in a chain of reasoning about causal mechanisms (Goldthorpe 1994).

In this paper, we first discuss the role of time in causal inferences in the social sciences. This will help us to recast a number of research problems so that fundamental design issues can be addressed more directly. Then, based on various examples, we will describe in detail how different observation plans (cross-secional, panel, and event history oriented) affect causal analysis. Although longitudinal data are no panacea, they are obviously more effective in causal analysis and have less inferential limitations.

We will limit our attention to data generated by continuous-time, discrete-state substantive processes. These processes have been characterized by James Coleman (1981:6) in the following general way: (1) there are units - which may be individuals, organizations, societies, or whatever - that change from one discrete state to another; (2) these changes (or events) can occur at any point in time and are not restricted to predetermined points in time; and (3) there are time-constant and/or time-dependent factors influencing the events. Examples of such process type can be given for a wide variety of social research fields (see, e.g., Blossfeld, Hamerle, and Mayer 1989; Blossfeld and Rohwer 1995). Consider, for example, a person's job career. The job history may be described as the time spent in the first job and the date the person entered into this spell, the type and duration in the following job or the kind of non-employment and its duration, and so on.

2. Causal Statements in the Social Sciences

In this section, we will focus on the general role of time in causal inferences and also show that the idea of a causal effect can be represented as a change in the transition rate, if the dependent variable is discrete and can change its state at any time.

Correlation and Causation

To begin with, statements about causation should be distinguished from statements about association. In making correlational inferences, one can be satisfied to observe how the values of one variable are associated with the values of other variables over the population under study and perhaps over time. In this context, time is only important insofar as it determines the population under analysis or specifies the operational meaning of a particular variable (Holland 1986). Statements about associations describe what has happened. They are quite different from causal statements and are designed to give information about how events are produced or conditioned by other events.

Sometimes social scientists argue that because the units of sociological analysis continuously learn and change and involve actors with goals and beliefs, sociology can at best only provide systematic descriptions of phenomena at various points in history. This position is based on the view that causal statements about events are only possible if they are regulated by "eternal," time-less laws (Kelly and McGrath 1988). Of course, the assumption that such laws can be established with regard to social processes can reasonably be disputed. However, we are not forced to accept a simple contrast: either describing contingent events or assuming "eternal" laws. Many social phenomena show systematic temporal variations and patterned regularities under specific conditions that themselves are a legitimate focus of our efforts to understand social change (Kelly and McGrath 1988). Thus, sociology can do more than just describe the social world. This paper therefore emphasizes the usefulness of techniques of event history modeling as "new" approaches to the investigation of causal explanations.[1]

Causal Mechanisms and Substantive Theory

The identification of causal mechanisms has been one of the classic concerns in sociology. Causal statements are made to explain the occurrence of events, to understand why particular events happen, and to make predictions when the situation changes (Marini and Singer 1988). Although sociologists sometimes seem to be

---

[1] We speak of a "new" approach just to emphasize the contrast to traditional "causal analysis" based on structural equation models which are basically time-less models. See the discussion in Bollen (1989), Campbell, Mutran and Parker (1987), or Faulbaum and Bentler (1994).

opposed to using the word "cause," they are far less reluctant to apply very similar words like "force," "agency," or "control," when trying to understand social phenomena.

There is consensus in the fact that causal inferences cannot simply and directly be made from empirical data, regardless of whether they are collected through ingenious research designs or summarized by particularly advanced statistical models. Thus, using event history observation plans and event history analysis models per se will not allow us to prove causality, as is the case for all other statistical techniques. However, as we will see in the next section, event-oriented observation designs offer richer information and, as we will try to demonstrate in this paper, event history models provide more appropriate techniques for exploring causal relations.

It seems useful to treat causality as being a property of theoretical statements rather than the empirical world itself (Goldthorpe 1994). In sociology, causal statements are based primarily on substantive hypotheses which the researcher develops about the social world. In this sense, causal inference is theoretically driven (Freedman 1991) and it will always reflect the changing state of sociological knowledge in a field.[2] Of course, descriptive statements are also dependent on theoretical views guiding the selection processes and providing the categories underlying every description. The crucial point is however that causal statements need a theoretical argument specifying the particular mechanism of how a cause produces an effect or, more generally, in which way interdependent forces affect each other in a given setting over time.

Therefore, the important task of event history modeling is not to demonstrate causal processes directly, but to establish relevant empirical evidence that can serve as a link in a chain of reasoning about causal mechanisms (Goldthorpe 1994). In this respect, event history models might be particularly helpful instruments because they allow a <u>time-related empirical representation of the structure of causal arguments</u>.

<u>Attributes, Causes and Time-Constant Variables</u>

Holland (1986) tried to establish some links between causal inference and statistical modeling. In particular, he emphasized that for a conception of causality it is essential that each unit of a population must be exposable to any of the various levels of a cause, at least hypothetically. He argues, for example, that the schooling a student

---

[2] Causal relations are always identified against the background of some field, and specification of a field is critical to the identification of an observed relation (Marini and Singer 1988).

receives can be a cause of the student's performance on a test, whereas the student's race or sex cannot. In the former case it seems possible to contemplate measuring the causal effect, whereas in the latter cases, where we have the enduring attributes of a student, all that can be discussed is association (see also Yamaguchi 1991).

We agree with Holland that it is essential for causal statements to imply counterfactual reasoning: if the cause had been different, there would have been another outcome, at least with a certain probability. In this sense, counterfactual statements reflect imagined situations. It is not always clear, however, which characteristics of a situation can sensibly be assumed to be variable, i.e. can be used in counterfactual reasoning, and which characteristics should be regarded as fixed. At least to some degree, the distinction depends on the field of investigation. For example, from a sociological point of view what is important with regard to sex is not the biological attributes per se, but the social meaning attached to these attributes. The social meaning of these attributes can change regardless of whether their biological basis changes or not. For example, societal rules might change to create more equality between the races or sexes. We therefore think that in sociological applications counterfactuals can also be meaningfully applied to such attributes. They can be represented as time-constant "variables" in statistical models to investigate their possible impact on some outcome to be explained. It is, however, important to be quite explicit about the sociological meaning of causal statements which involve references to biological or ethnic attributes. There is, for example, no eternal law connecting gender and/or race with wage differentials. But probably there are social mechanisms which connect gender and ethnic differences with different opportunities in the labor market.

Causes and Time-Dependent Variables

The meaning of the counterfactual reasoning of causal statements is that causes are states which could be different from what they actually are. However, the consequences of conditions that could be different from their actual state are obviously not observable.[3] It means that it is simply impossible to observe the effect that would have happened on the same unit of analysis, if it were exposed to another condition at the same time. To find an empirical approach to causal statements, the researcher must look at conditions which actually do change in time. These changes are events. More formally, an event is a change in a variable, and this change must happen at a specific point in time. This implies that the most obvious empirical representation of causes is

[3] Holland (1986) calls this "the fundamental problem of causal inference."

in terms of variables that can change their states over time. This statement is linked very naturally with the concept of time-dependent covariates in event history analysis. The role of a time-dependent covariate in this approach is to indicate that a (qualitative or metric) causal factor has changed its state at a specific time and that the unit under study is exposed to another causal condition. For example, in the case of gender the causal events might be the steps in the acquisition of gender roles over the life course or the exposure to sex-specific opportunities in the labor market at a specific historical time. Thus, a time-constant variable "gender" should ideally be replaced in an empirical analysis by time-changing events assumed to produce sex-specific differences in the life history of men and women. Of course, in empirical research that is not always possible, so that one very often has to rely on time-constant "variables" as well. However, it is important to recognize that for these variables the implied longitudinal causal relation is not examined. For example, if we observe an association amongst people with different levels of educational attainment and their job opportunities, then we can normally draw the conclusion that changes in job opportunities are a result of changes in educational attainment level. The implied idea is the following: if we started having people with the lowest educational attainment level and followed them over the life course, they would presumably differ in their rates to attaining higher levels of educational attainment and this would produce changes in job opportunities. Whether this would be the case for each individual is not very clear from a study that is based on people with different levels of educational attainment. In particular, one would expect that the causal relationship between education and job opportunities would radically be altered if all people acquired a higher (or the highest) level of educational attainment.[4] Thus, the two statements - the first about associations across different members of a population and the second about dependencies in the life course for each individual member of the population - are quite different; one type of statement can be empirically true while the other one can be empirically false. Therefore, statements of the first type cannot be regarded as substitutes for statements of the second type. However, since all causal propositions have consequences for longitudinal change (see Lieberson 1985), only time-changing variables provide the most convincing empirical evidence of causal relations.[5]

---

[4] A longitudinal approach would provide, however, the opportunity to study these kinds of changes in the causal relationships over time.

[5] There is also another aspect that is important here (see Lieberson 1985): causal relationships can be symmetric or asymmetric. In examining the causal influence of a change in a variable X on a change in a dependent variable Y, one has to consider whether shifts to a given value of X from either direction have the same consequences for Y. For example, rarely do researchers consider whether an upward shift on the prestige scale, say from 20 to 40, will lead to a different

Time Order and Causal Effects

We can summarize our view of causal statements in the following way:

$$\Delta X_t \rightarrow \Delta Y_{t'} \qquad t < t'$$

meaning that a change in variable X_t at time t is a cause of a change in variable $Y_{t'}$ at a later point in time, t'. It is not implied, of course, that X_t is the only cause which might affect $Y_{t'}$. So we should speak of causal conditions to stress that there might be, and normally is, a quite complex set of causes.[6] Thus, if causal statements are studied empirically, they must intrinsically be related to time. There are three important aspects. First, to speak of a change in variables necessarily implies reference to a time axis. We need at least two points in time to observe that a variable has changed its value. Of course, at least approximately, we can say that a variable has changed its value at a specific point in time.[7] Therefore, we use the symbols to refer to changes in the values of the time-dependent variable ΔX_t and the state variable ΔY_t at time t. This leads to the important point that causal statements relate changes in two (or more) variables.

---

outcome of Y (say family decisions) than would a downward shift of X from 60 to 40. In other words, most researchers assume symmetry. However, even if a change is reversible, the causal process may not be. The question is: if a change in a variable X causes a change in another one, Y, what happens to Y if X returns to its earlier level? "Assuming everything else is constant, a process is reversible, if the level of Y also returns to its initial condition; a process is irreversible if Y does not return to its earlier level. Observe that it is the process - not the event - that is being described as reversible or irreversible." (Lieberson 1985:66)

6 It is important to note here that the effect of a variable X is always measured relative to other causes. A conjunctive plurality of causes occurs if various factors must be jointly present to produce an effect. Disjunctive plurality of causes, on the other hand, occurs if the effect is produced by each of several factors alone, and the joint occurrence of two or more factors does not alter the effect (see the extensive discussion in Marini and Singer (1988)).

7 Statements like this implicitly refer to some specification of "point in time." The meaning normally depends on the kind of events which are to be described, for instance, a marriage, the birth of a child, or to become unemployed. In event history text books, normally a continuous time axis for purposes of mathematical modeling is assumed (see Blossfeld and Rowher 1995). This should however be understood as an idealized way of representing social time. Here we are using mathematical concepts to speak about social reality, so we will disregard the dispute about whether time is "continuous" (in the mathematical sense of this word), or not.

Second, there is a time ordering between causes and effects. The cause must precede the effect in time: $t < t'$, in the formal representation given above. This seems to be generally accepted.[8] As an implication, there must be a temporal interval between the change in a variable representing a cause, and a change in the variable representing a corresponding effect. It is important to realize that the role of time in causal explanations does not only lie in specifying a temporal order in which the effect follows the cause in time. It additionally implies that a temporal interval is necessary for the cause to have an impact (Kelly and McGrath 1988). In other words, if the cause has to precede the effect in time, it takes some finite amount of time for the cause to produce the effect. The time interval may be very short or very long, but can never be zero or infinity (Kelly and McGrath 1988). Some effects take place almost instantaneously. For example, if the effect occurs at microsecond intervals, then the process must be observed in these small time units to uncover causal relations. However, some effects may occur in a time interval too small to be measured by any given methods, so that cause and effect seem to occurat the same point in time. Apparent simultaneity is often the case in those social science applications where basic observation intervals are relatively crude (e.g. days, months, or even years), such as, for example, yearly data about first marriage and first childbirth (Blossfeld, Manting and Rohwer 1993). For these parallel processes, the events "first marriage" and "first childbirth" may be functionally interdependent, but whether these two events are observed simultaneously or successively depends on the degree of temporal refinement of the scale used in making the observations. Other effects need a long time until they start to occur. Thus, there is a delay or lag between cause and effect that must be specified in an appropriate causal analysis. However, in most of the current sociological theories and interpretations of research findings this interval is left unspecified.

This immediately leads to a third point. In addition to the question of how long the delay between the timing of the cause and the beginning of the unfolding of the effect is, there might be different shapes of how the causal effect Y_t unfolds over time. While the problem of time-lags is widely recognized in social science literature, there is almost no information with respect to the temporal shapes of effects (Kelly and McGrath 1988). Social scientists seem to be quite ignorant with respect to the fact that causal effects could be highly time-dependent, too. The panels of Figure 1 illustrate several possible shapes these effects may trace over time. In Figure 1a, there is an almost all-at-once change that is then maintained; in Figure 1b, the effect occurs with some lengthy time-lag and is then time-invariant; in Figure 1c, the effect starts almost

[8] See, for instance, the discussion in Eells (1991, Ch.,5).

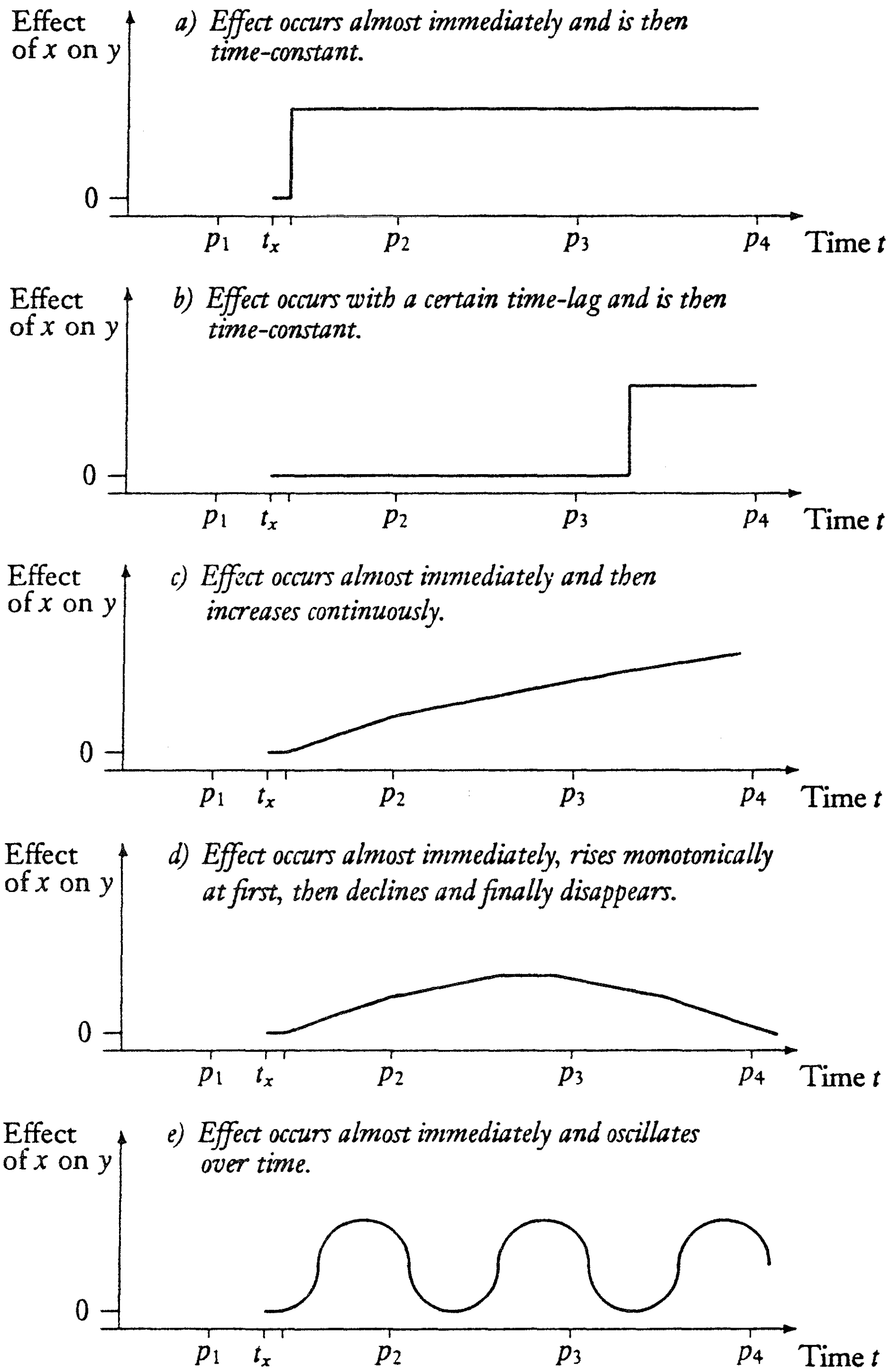

Figure 1 Different temporal shapes of how a change in a variable x, occurring at point in time t_x, effects a change in a variable y.

immediately and then gradually increases; in Figure 1d, there is an almost all-at-once increase, which reaches a maximum after some time and then decreases; finally, in Figure 1e, a cyclical effect pattern over time is described. Thus, an appropriate understanding of causal relations between variables should take into account that the causal relationship itself may change over time. This seems particularly important in sociological applications of causal reasoning. In these applications we generally cannot rely on the assumption of eternal, time-less laws but have to recognize that the causal mechanisms may change during the development of social processes.

Actors and Probabilistic Causal Relations

It seems agreed that social phenomena are always directly or indirectly based on actions of individuals. This clearly separates the social from the natural sciences. Sociology therefore does not deal with associations among variables per se, but with variables that are associated via acting people. There are at least three consequences for causal relations: First, in methodological terms, this means that if individuals relate causes and effects through their actions, then research on social processes should at best be based on individual longitudinal data (Coleman and Hao 1989; Coleman 1990). This is why life history data on individuals, and not aggregated longitudinal data, provide the most appropriate information for the analyses of social processes. Only with these data can one trace the courses of action at the level of each individual over time. Second, in theoretical terms it means that the explaining or understanding of social processes requires a time-related specification of (1) the past and present conditions under which people act,[9] (2) the many and possibly conflicting goals that they pursue at the present time, (3) the beliefs and expectations guiding the behavior, and (4) the actions that probably will follow in the future.[10]

---

[9] These conditions are, of course, heavily molded by social structural regularities in the past and the present. Sociology must always be a historical discipline (Goldthorpe 1991).

[10] Sometimes it is argued that, since human actors act intentionally and behavior is goal-oriented, the intentions or motives of actors to bring about some effect in the future causes the actor to behave in a specific way in the present (Marini and Singer 1988). This does not however contradict a causal view. One simply has to distinguish intentions, motives or plans as they occur in the present from their impact on the behavior which follows their formation temporally, and from the final result, as an outcome of the behavior. An expectation about a future state of affairs should clearly be distinguished from what eventually happens in the future. Therefore, the fact that social agents can behave intentionally, based on expectations, does not reverse the time order underlying our causal statements.

Third, if it is people that are doing the acting, then causal inference must also take into account the free will of individuals. This introduces an essential element of indeterminacy into causal inferences. This means that in sociology we can only reasonably account for and model the generality but not the determinacy of behavior. The aim of substantive and statistical models must therefore be to capture common elements in the behavior of people, or patterns of action that recur in many cases (Goldthorpe 1994). This means that in sociological applications randomness has to enter as a defining characteristic of causal models. We can only hope to make sensible causal statements about how a given or (hypothesized) change in variable X_t in the past affects the probability of a change in variable $Y_{t'}$ in the future. Correspondingly, the basic causal relation becomes

$$\Delta X_t \rightarrow \Delta \Pr(\Delta Y_{t'}) \qquad t < t' \qquad (1.1)$$

This means that a change in the time-dependent covariate X_t will change the probability that the dependent variable $Y_{t'}$ will change in the future ($t' > t$). In sociology, this interpretation seems more appropriate than the traditional deterministic approach. The essential difference is not that our knowledge about causes is insufficient because it only allows probabilistic statements, but that the causal effect to be explained is a probability. Thus, probability in this context is not just a technical term anymore, but is considered as a theoretical one: it is the propensity of social agents to change their behavior.

Causal Statements and Limited Empirical Observations

A quite different type of randomness related to making inferences occurs if causal statements are applied to real-world situations in the social sciences. There are at least four additional reasons to expect further randomness in empirical studies. These are basically the same ones that occur in deterministic approaches and are well-known from traditional regression modeling (Lieberson 1991). The first one is measurement error, a serious problem in empirical social research, which means that the observed data deviate somewhat from the predicted pattern without invalidating the causal proposition. The second reason is particularly important in the case of non-experimental data. It is often the case that complex multivariate causal relations operate in the social world. Thus, a given outcome can occur because of the presence of more than one influencing factor. Moreover, it may also not occur at times because the impact of one independent variable is outweighed by other influences working in the opposite direction. In these situations, the observed influence of the cause is only approximate, unless one can control for the other important factors. The third motive is that sociologists often do not know or are not able to measure all of the important factors. Thus, social scientists have to relinquish the idea of a complete measurement of causal effects, even if they would like to make a deterministic proposition. Finally, sometimes chance affects observed outcomes in the social world. It is not important here to decide whether chance per se exists or whether it is only a surrogate for the poor state of our knowledge of additional influences and/or inadequate measurement.

In summary, these problems imply that social scientists can only hope to make empirical statements with a probabilistic character. This situation normally leads to identification problems. Without strong assumptions about missing information and errors in the available data, it is generally not possible to find definite statements about causal relations (see Blossfeld and Rohwer 1995).

A Simplistic Conception of Causal Relations

At this point it is important to stress that the concept of causal relation is a rather special abstraction implying a time-related structure that does not immediately follow from our sensory impressions. Consider the following example in Figure 2 where we characterize the necessary time-related observations of a unit being affected by a causal effect. This figure shows that an empirical representation of the most simple causal effect, (i.e. (1) where the condition X_t changes (from one state $X_{t1}= x_1$ to another one $X_{t2} = x_2$) and (2) is then constant afterwards, (3) the change in Y_t (from $Y_{t2} = y_1$ to $Y_{t3} = y_2$) takes place almost instantaneously and (4) is then also time-constant afterwards) needs at least three points in time where the researcher must note the states of the independent and dependent variables, respectively.[11] This is because, if we assume that a change in the independent variable X_t has taken place at t_2, then, to be able to fix the particular change in the condition in the past, we need to know the state of the independent variable X_t at an earlier time, t_1 (see Figure 2). For the dependent variable Y_t we need an observation before the effect has started to occur. Assuming everything else is constant, this observation can be made, at the latest at point t_2 because the effect has to follow the cause in time. To evaluate whether the hypothesized effect has indeed taken place at a later time, t_3, we must again note the state of the dependent variable Y_t. Thus, a simplistic representation of a causal effect exists when we compare the change in the observations for the independent variable in the past and the present with the change in the observations

[11] This example is instructive because Lazarsfeld (1948,1972), and many others after him have argued that for causal inferences two observations of the units would be sufficient.

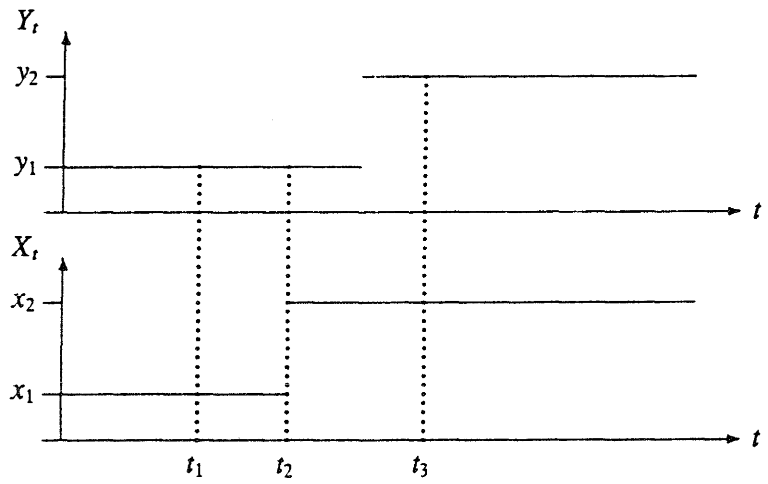

Figure 2 Observation of a simple causal effect.

for the dependent variable in the present and in the future and link both changes with a substantive argument.[12]

However, this is only a simple and fairly unrealistic example of a causal effect. In the case of observational data in the social sciences, where there are many (qualitative and metric) causal variables that might change their values at any point in time, when their causal effects might have various delays and different shapes in time (see Figure 1), then the quantity of the observed causal effect as shown in Figure 1 will strongly depend on when the measurements at the three points in time are taken.

Thus, what can we say about the causal effect(s) at any given point in time if the situation is more complex? A paradox occurs: the concept of causal effect depends intrinsically on comparisons between changes in both the independent and dependent variables in at least three points in time. Yet the concept of causal effect should itself reflect a state of a unit of observation at any point in time as being an appropriate one in real empirical situations. Thus, what is still needed in our discussion is a concept that represents the quantity of the causal effect at any point in time.

Causal Effects and Changes in Transition Rates

If the dependent variable is discrete and can change its state at any time, then the transition rate framework offers a time-point-related representation for the causal effect. We briefly want to develop this idea.

[12] Indeed such a simplistic idea of the causal effect is the basis of all panel designs, as shown in the next section.

Let us first start with the dependent variable, Y_t, and its changes in the future (as a consequence of a change in a causal factor). In particular, we are interested in changes of states occupied by the units of analysis. The state space is assumed to be discrete, and so the possible changes are discrete. We assume that a unit enters at time t_0 into the (origin) state j, that is $Y_{t0} = j$. The basic form of change to be explained in the transition rate framework is the probability of a change in Y_t from an origin state j to a destination state k (while $t > t_0$). Now, we need a concept that allows describing the development of the process at every point in time, while the process is going on, and that, for its definition, only relies on information about the past development of the process. The crucial concept that can be used for this purpose is the transition rate. To define this concept, let us first introduce a random variable T to represent the duration, beginning at t_0, until a change in the dependent variable, that is a transition from (origin) state j to (destination) state k, occurs. To simplify the notation we will assume that $t_0 = 0$. Then, the following probability can be defined:

$$\Pr(t \leq t < t' \mid T \geq t) \qquad t < t' \qquad (1.2)$$

This is the probability that an event occurs in the time interval from t to t', given that no event (transition) has occurred before, that is, in the interval from 0 to t. This probability is well defined and obviously well suited to describe the temporal evolution of the process. The definition refers to each point in time while the process is evolving, and thereby can express the idea of change during its development. Also, the definition only relies on information about the past of the process, what has happened up to the present point in time, t. Therefore, the concept defined in (1) can sensibly be used to describe the process before it has finished for all individuals in the population. Assume that we know the probabilities defined in (1.2) for all points in time up to a certain point t^*. Then we have a description of the process up to this point, and this description is possible without knowing how the process will develop in the future, i.e. for $t > t^*$.

Since our mathematical model is based on a continuous time axis, one can in the expression (1.2) let t'- t approach zero. However, as the length of the time interval approaches zero, the concept of change in the dependent variable would simply disappear because the probability that a change takes place in an interval of zero length is zero:

$$\lim_{t' \to t} \Pr(t \leq t < t' \mid T \geq t) = 0$$

To avoid this, we regard the ratio of the transition probability to the length of the time interval to represent the probability of future changes in the dependent variable per unit of time (Coleman 1968), i.e. we consider

$$\Pr(t \leq t < t' \mid T \geq t) \,/\, (t' - t)$$

This allows us to define the limit

$$r(t) = \lim_{t' \to t} \Pr(t \leq t < t' \mid T \geq t) \,/\, (t' - t)$$

and we arrive at the central concept of the transition rate. This concept obviously provides the possibility of giving a local, time-related description of how the process (defined by a single episode) evolves over time. We can interpret r(t) as the propensity to change the state, from origin j to destination k, at t. But one should note that this propensity is defined in relation to a risk set, the risk set at t, i.e. the set of individuals who can experience the event because they have not already had the event before t. Having introduced the basic concept of a transition rate, we can finally formulate our basic modeling approach. The preliminary description in (1.1) can now be restated in a somewhat more precise form as

$$r(t) = g(t,x) \qquad (1.4)$$

This is the basic form of a transition rate model. The central idea is to make the transition rate, which describes a process evolving in time, dependent on time and on a set of covariates, x. Obviously, we also need the "variable" time (t) on the right-hand side of the model equation. However, it must be stressed that a sensible causal relation can only be assumed for the dependency of the transition rate on the covariates. The causal reasoning underlying the modeling approach (1.4) is

$$\Delta X_t \rightarrow \Delta r(t') \qquad t < t'$$

As a causal effect, the changes in some covariates in the past may lead to changes in the transition rate in the future, which in turn describe the propensity that the units under study will change in some presupposed state space. As discussed above, this causal interpretation requires that we take the temporal order in which the process evolves very seriously. At any given point in time, t, the transition rate r(t) can be made dependent on conditions that happened to occur in the past, i.e. before t, but not on what is the case at t or in the future after t. With respect to these individuals, and while the process is evolving, there is always a distinction in past, present, and future. This is particularly important for a causal view of the process. The past conditions the present, and what happens in the present shapes the future. There are many possibilities to specify the functional relationship g(.) in (1.4) (see Blossfeld and Rohwer 1995).

It is sometimes argued that sociologists should give up the causal analytical point of view in favor of a systems view because the operation of causal forces is mutually interdependent and variables change each other more or less simultaneously in many systems (Marini and Singer 1988). However, even in systems of interdependent processes time does not run backwards, and change in one of the interdependent variables will take (at least a small amount of) time to produce a change in another one. Thus, in systems of variables there cannot be any simultaneity of causes and their effects. This allows to demonstrate that a causal approach to interdependent systems is possible with the help of the transition rate concept (see Blossfeld and Rohwer 1995). In other words, the systems view is not a substitute for a proper causal approach in our field (Kelly and McGrath 1988).

Since the transition rate is indeed an abstraction, it is necessary to relate it back to quantities that are directly observable, that is to frequencies of state occupancies at particular points in time. To support such inferences, some additional statistical concepts are useful which are extensively described by Blossfeld and Rohwer (1995).

3. Causal Modeling and Observation Plans

Over the last 20 years, social scientists have been collecting event history data with increasing frequency. This is not an accidental trend, nor does it reflect a prevailing type of fashion in survey research. Instead, as discussed in the previous section, it indicates a growing recognition among social scientists that event history data is often the most appropriate empirical information one can get for a causal analysis.

To collect data generated by a continuous-time, discrete-state substantive process, different observation plans have been used (Coleman 1981; Tuma and Hannan 1984). With regard to the extent of detail about the process of change, one can distinguish between cross-sectional data, panel data, event count data, event sequence data, and event history data.

In this paper, we will not discuss event count data (see Barron 1993), which simply record the number of different types of events for each unit (e.g. the number of upward, downward, or lateral moves in the employment career in a period of 10 years), and event sequence data, which document the sequence of states occupied by each unit, as they are rarely used in the social sciences. On the other hand, cross-sectional data and panel data are standard sociological data types (Tuma and Hannan 1984). It is, therefore, particularly intriguing to compare event history data with cross-sectional and panel data. We will use the example shown in Figure 3. In this figure, an individual's family career is observed in a cross-sectional survey, a panel survey and an event-oriented survey.

Cross-Sectional Data

Let us first discuss the cross-sectional observation. In the social sciences, this is the most common form of data for assessing the determinants of behavior. The family history of the individual in Figure 3 is represented by one single point in time: his or her marital state at the time of interview. Thus, a cross-sectional sample is only a "snapshot" of the substantive process being studied. The point in time when researchers take that "picture" is normally not determined by hypotheses about the dynamics of the substantive process itself, but by external considerations like getting research funds, finding an appropriate institute to conduct the survey etc.

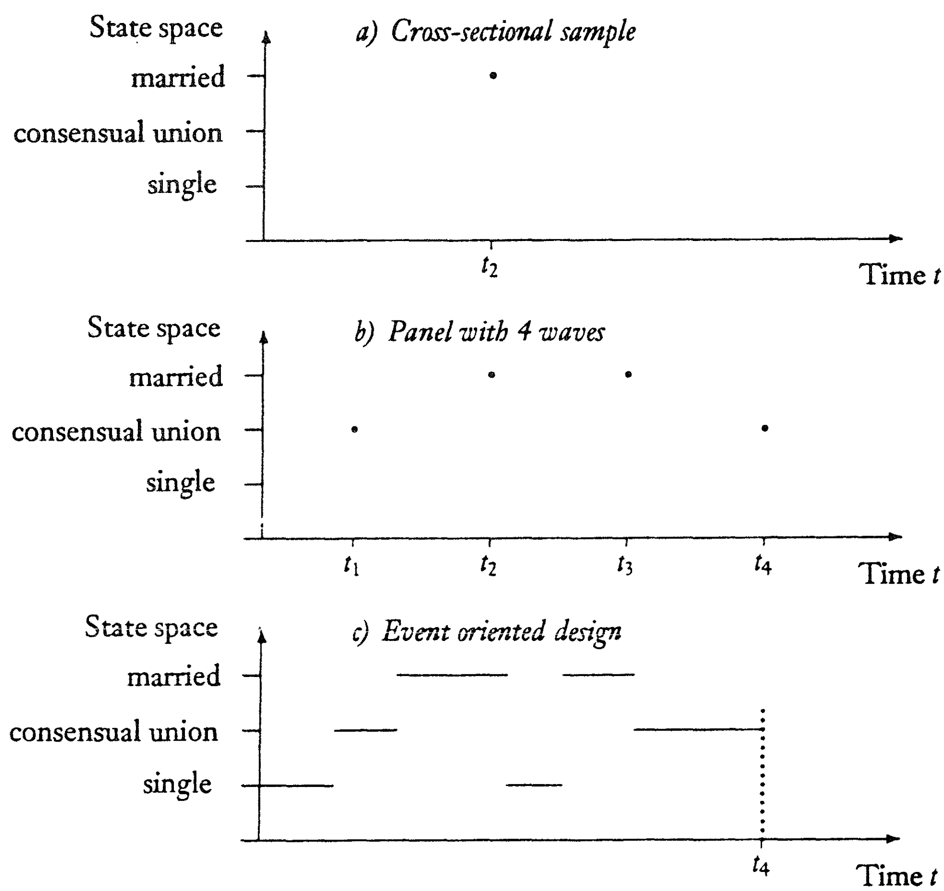

Figure 3 Observation of an individual's family career on the basis of a cross-sectional survey, a panel study and an event history oriented design.

Coleman (1981) has demonstrated that one must be cautious in drawing inferences about explanatory variables on the basis of such data because, implicitly or explicitly, social researchers must assume that the substantive process under study is in some kind of statistical equilibrium. Statistical equilibrium or stability of the process means that whilst individuals (or any other unit of analysis) may change their states over time, the state probabilities are fairly trendless or stable. Therefore, an equilibrium of the process requires that the inflows to and the outflows from each of the discrete states be equal over time to a large extent. It is only in such cases that one can reasonably assess the estimates of logit and log-linear analyses, as shown by Coleman (1981).

However, these estimates are ambiguous because they only represent the net differences in the effects of independent variables (Coleman 1981; Blossfeld and Rohwer 1995). If there is no equilibrium in the process, cross-sectional coefficients may not only be ambiguous, but even present a completely misleading picture. In a recent study on unemployment incidence, Rosenthal (1991), for example, demonstrates how confusing cross-sectional estimates can be if the unemployment rate increases or decreases in a specific region and if the process of change is, therefore, not in equilibrium.

In the social sciences one can expect that stability is very rare. For example, life history studies (Mayer 1990; Blossfeld 1989, 1995a, 1995b) show that change across age, cohort and historical period is an enduring and important feature in all domains of modern individuals' lives (Mayer and Tuma 1990); organizational studies demonstrate that most social organizations seem to follow a program of growth and not of stability; and most modern societies reveal an accelerating rate of change in almost all of their subsystems (cf. the rapid changes in family systems, job structures, educational systems etc., see Heinz 1991a, 1991b, 1992; Mayer 1990; Blossfeld 1989, 1995a, 1995b). But even in areas considered to be fairly stable, one must ask the crucial methodological question: to what extent is the process under study close to an equilibrium (Tuma and Hannan 1984)? This question can only be answered if longitudinal data are applied, as longitudinal data are the only type of data which indicate whether a steady state actually exists, or how long it will take until a system returns to a new equilibrium after some external upheaval.

Beyond the crucial assumption of process stability, cross-sectional data have many inferential limitations with regard to causal modeling. We want to address at least some of the more important problems here.[13]

Direction of causality. There are only a few situations in which the direction of causality can be established based on cross-sectional data (Davies 1987). For example, consider the strong positive association between parental socioeconomic characteristics and educational attainment of sons and daughters, controlling for other important influences (Shavit and Blossfeld 1993). A convincing interpretation of this effect might be that being born into a middle class family increases the likelihood of attaining a university degree because one is unable to think of any other plausible explanation for the statistical association. However, such recursive relationships, in which all the causal linkages run "one way" and have no "feedback" effects, are rare in social science research. For example, there is very often an association between the age of the youngest child and female labor force participation in modern industrialized societies (Blossfeld and Rohwer 1995). The common interpretation is that there is a one-way causality with young children tending to keep mothers at home. However, it is quite possible that the lack of jobs encourages women to enter into marriage and motherhood, suggesting a reversed relationship (Davies 1987).

The ambiguity of causation seems to be particularly important for the modeling of the relationship between attitudes and behavior. There are two interesting aspects of this relationship: there is a direct effect in which behavior affects attitudes, and there is a "feedback" process where attitudes change behavior (Davies 1987). Footnote: The relationship between attitudes and behavior suggests that there is some kind of inertia (or positive feedback) which means that the probability of a specific behavior increases as a monotonic function of attitudes and attitudes depend on previous behavior (Davies and Crouchley 1985). The well-known disputes among sociologists, as to whether value change engenders change in social structure, or whether structural change leads to changing values of individuals, often originates from the fact that cross-sectional surveys can only assess the net association of these two processes.

Various strengths of reciprocal effects. Connected with the inability of establishing the direction of causality in cross-sectional surveys is the drawback that these data cannot be used to discover the different strengths of reciprocal effects. For example, many demographic studies have shown that first marriage and first motherhood are closely interrelated (Blossfeld and Huinink 1991). To understand what has been happening

[13] These problems are however not mutually exclusive.

with regard to family formation in modern societies, it might be of interest not only to know the effect of marriage on birth rates, but also the effect of pregnancy or first birth on getting married (Blossfeld and Huinink 1991; Blossfeld 1995a); and perhaps, how these effects have changed over historical time (Manting 1994).

Observational data. Most sociological research is based on nonexperimental observations of social processes and these processes are highly selective. For example, Lieberson (1985) in a study examining the influence of type of school (private vs. public) on test performance among students distinguished at least three types of non-random processes: (1) there is self-selectivity, in which the units of analysis sort themselves out by choice (e.g. specific students choose specific types of schools); (2) there is selective assignment by the independent variable itself, which determines, say, what members of a population are exposed to specific levels of the independent variable (e.g. schools select their students based on their past achievement); and (3) there is also selectivity due to forces exogenous to variables under consideration at the time (socioeconomic background, ethnicity, gender, previous school career, changes of intelligence over age etc.); and many of these sources are not only not observed, but effectively unmeasurable. Although no longitudinal study will be able to overcome all the problems of identification of causal effects, cross-sectional data offer the worst of all opportunities to disentangle the effects of the causal factors of interest on the outcome from other forces operating at the same time because these data are least informative about the process of change. Cross-sectional analysis therefore requires a particularly careful justification, and the results must always be appropriately qualified (Davies 1987; Pickles and Davies 1989).

Previous history. There is one aspect of observational data that deserves special attention in the social sciences. Life courses of individuals (and other units of analysis like organizations etc.) involve complex and cumulative time-related layers of selectivity (Mayer 1991). Therefore, there is a strong likelihood that specific individuals have been entering a specific origin state. In particular, life course research has shown that the past is an indispensible factor in understanding the present (Heinz 1991a, 1991b, 1992; Mayer 1990). Cross-sectional analysis may be performed with some proxy-variables as well as with assumptions of the causal order and interdependencies between the various explanatory variables. However, it is often not possible to appropriately trace back the time-related selective processes operating in the previous history, as these data are simply not available. Thus, the normal control approaches in cross-sectional statistical techniques will rarely be successful in isolating the influence of some specific causal force (Lieberson 1985).

Age and cohort effects. Cross-sectional data cannot be used to distinguish age and cohort effects (Tuma and Hannan 1984; Davies 1987). However, in many social science applications it is of substantive importance to know whether the behavior of people (e.g. their tendency to vote for a specific party) is different because they belong to different age groups or whether they are members of different birth cohorts (Blossfeld 1986, 1989).

Historical settings. Cross-sectional data are not able to take into account the fact that processes emerge in particular historical settings. For example, in addition to individual resources (age, education, labor force experience etc.), there are at least two ways in which a changing labor market structure affects career opportunities. The first is that people start their careers in different structural contexts. It has often been assumed that these specific historic conditions at the point of entry into the labor market have a substantial impact upon people's subsequent careers. This kind of influence is generally called a cohort effect (Glenn 1977). The second way that changing labor market structure influences career opportunities is that it improves or worsens the career prospects of all people within the labor market at a given time. For example, in a favorable economic situation with low unemployment, there will be a relatively wide range of opportunities. This kind of influence is generally called a period effect (Mason and Fienberg 1985). With longitudinal data, Blossfeld (1986) has shown that lifecourse, cohort, and period effects can be identified based on substantively developed measures of these concepts (Rodgers 1982), and that these effects represent central mechanisms of career mobility that must be distinguished.

Multiple clocks, historical eras and point-in-time events. From a theoretical or conceptual point of view, multiple clocks, historical eras and point-in-time events very often influence the substantive process being studied (Mayer and Tuma 1990). For example, in demographic studies of divorce, types of clocks, such as age of respondent, time of cohabitation, duration of marriage, ages of children, as well as different phases in the state of the business cycle, or changes in national (divorce) laws are of importance (Blossfeld, DeRose, Hoem, and Rohwer 1993). With respect to cross-sectional data, such relationships can hardly be studied without making strong untestable assumptions.

Contextual processes at different levels. Social scientists are very often interested in the influences of contextual processes at different aggregation levels (Huinink 1989). Contextual process effects refer to situations where changes in the group contexts themselves influence the dependent variable. For example, career mobility of an individual may be conceptualized as being dependent on changes in resources at the individual level (e.g. social background, educational attainment, experience etc.), the success of the firm in which he/she is employed (e.g. expansion or contraction of the organization) at the intermediate level, and changes in the business cycle at the macro level (Blossfeld 1986; DiPrete 1993). Cross-sectional data do not provide an adequate opportunity for the study of such influences at different levels (Mayer and Tuma 1990).

Duration dependence. Another problem of cross-sectional data is that they are inherently ambiguous with respect to their interpretation at the level of the unit of observation. Suppose we know that in the Federal Republic of Germany, 30.6 per cent of employed women were working part-time in 1970 (Blossfeld and Rohwer 1995). At the one extreme, this might be interpreted to imply that each employed woman had a 30.6 per cent chance of being employed part-time in this year, but on the other, one could infer that 30.6 per cent of the employed women always worked part-time and 69.4 per cent were full-timers only. In other words, cross-sectional data do not convey information about the time women spent in these different employment forms. They are therefore open to quite different substantive interpretations (Heckman and Willis 1977; Flinn and Heckman 1982). In the first case, each woman would be expected to move back and forth between part-time and full-time employment. In the second, there is no mobility between part-time and full-time work, and the estimated percentages describe the proportions of two completely different groups of employed women. From an analytical point of view, it is therefore important to have data about durations in a state. Also, repeated cross-sectional analysis using comparable samples of the same population (e.g. a series of microcensuses or cross-sectional surveys), can only show net change, not the flow of individuals.

Variability in state dependencies. In many situations cross-sectional data are problematic because the rate of change is strongly state dependent and entries into and exits from these states are highly variable over time (e.g. over the life course and historical period or across cohorts). For example, it is well-known that the roles of wives and mothers (the latter in particular) have been central in women's lives. Therefore, the family cycle concept has frequently been used in sociology to describe significant changes in the circumstances which affect the availability of women for paid work outside the home. The basic idea is that there is a set of ordered stages primarily defined by variations in family composition and size that could be described with cross-sectional data. However, this view often leads to the tendency to assume that what happens to different women in various phases in the family cycle at one point in time is similar to the pattern that women experience when they make these transitions in different historical times (which has been called the "life course fallacy"). Moreover, there is the well-known problem that individuals and families often fail to conform to the assumption of a single progression through a given number of stages in a predetermined order. At least three reasons for this may exist (Murphy 1991): (1) the chronology of timing of events may not conform to the ideal model, for example childbearing may start before marriage; (2) many stages are not reached, for example, by never-married persons; and (3) the full set of stages may be truncated by events such as death or marital breakdown. Such complex constellations between the family cycle and women's labor force participation could hardly be meaningfully described or studied on the basis of cross-sectional data (see also Blossfeld 1995b).

Changes in outcomes. Cross-sectional models very often have a tendency to over-predict change and consistently over-estimate the importance of explanatory variables (Davies 1987). The reason for this phenomenon is that these analyses cannot be based on how changes in explanatory variables engender changes in outcomes. They are only concerned with how levels of explanatory variables "explain" an outcome at a specific point in time. However, if an outcome at time t (e.g. choice of mode of travel to work in June) is dependent on a previous outcome (e.g. established choice of mode of travel to work), and if both outcomes are positively influenced in the same way by an explanatory variable (e.g. merits of public transport), then the effect of the explanatory variable will reflect both the true positive influence of the explanatory variable on the outcome at time t and a positive spurious element due to that variable acting as a proxy for the omitted earlier outcome (established mode of travel to work). Thus, a cross-sectional analysis of the travel to work choice (e.g. public vs. private transport) would have a tendency to overpredict the effect of policy changes (e.g. fare increases or faster buses) because there is a strong behavioral inertia (Davies 1987).

These examples show that cross-sectional data have many severe inferential limitations for social scientists. Therefore, it is not surprising that causal conclusions based on cross-sectional data have often been radically altered after the processes were studied with longitudinal data (Lieberson 1985).

Longitudinal studies also have a much greater power than cross-sectional ones, both in the estimation of bias from missing data, and in the means for correcting it. This is because in longitudinal studies one often has data from previous points in time, thus enabling the characteristics of non-responders or lost units to be assessed with some precision. It is noteworthy that almost all the substantive knowledge concerning the biases associated with missing data, which all studies must seek to minimize, is derived from longitudinal studies (Medical Research Council 1992).

Although longitudinal data are no panacea, they are obviously more effective in causal analysis and have less inferential limitations (Magnusson, Bergmann and Torestad 1991; Arminger, Clogg and Sobel 1995; Clogg and Arminger 1993; Blossfeld 1995a, 1995b; Mayer 1990; Mayer and Tuma 1990). They are indispensable for the study of processes over the life course (of all types of units) and their relation to historical change. Therefore, research designs aimed at a causal understanding of social processes should be based on longitudinal data at the level of the units of analysis.

Panel Data

The temporal data most often available to sociologists are panel data, for which the same persons or units are re-interviewed or observed at a series of discrete points in time (Chamberlain 1984; Hsiao 1986; Arminger and Muller 1990; Engel and Reinecke 1994). Figure 3 shows four-wave panel in which the family career of the respondent was observed at four different points in time. This means that there is only information on states of the units at pre-determined survey points, but the course of the events between the survey points remains unknown.

Panel data normally contain more information than cross-sectional data, but involve well-known distortions created by the method itself (see, e.g., Magnusson and Bergmann 1990; Hunt 1985):

Panel bias. Respondents often answer many questions differently in the second and later waves than they did the first time; perhaps this is because they are less inhibited, or they mulled over or discussed the issues between questioning dates.

Modification of processes. Panels tend to influence the very phenomena they seek to observe - this changes the natural history of the processes being observed.

Regression to the mean. During the first panel wave, some panel members answer specific questions a bit differently than they normally would, due to chance circumstances at the time of measurement. For example, when a respondent has some problems during the first observation, he/she may perform poorly on some topics, but will report more accurately at the next panel wave. This phenomenon, known as "regression to the mean," can easily be mistaken by researchers for specific trends.

Attrition of the sample. In panel studies the composition of the sample normally diminishes selectively as time goes by. Therefore, what researchers observe in the panel may not give a true view of what has happened to their original sample.

Non-responses and missing data. In a cross-sectional analysis, one can afford to throw out a small number of cases with "non-responses" and "missing data," but in a long-term longitudinal study, throwing out incomplete cases at each round of observations can eventually leave a severely pruned sample having very different characteristics from the original one.

Fallacy of cohort centrism. Very often panel studies are focused on members of a specific cohort (cf., for example, the British National Child Study). In other words, these panels study respondents that were born in, grew up in, and have lived in a particular segment of history. There is therefore a danger that researchers might assume that what happens to a particular group of people over time reveals general principles of the life course (fallacy of cohort centrism). Many events may simply be specific for that generation.

Fallacy of period centrism. Many panel studies include just a few waves and, therefore, cover only a short period of historical time (cf. the German Socio-Economic Panel which now covers 10 years). At the time of a particular observation, special conditions may exist and this can result in an individual responding differently than he/she normally would (fallacy of historical period).

Confounded age, period and cohort effects. In any long-term panel study in sociology, three causal factors - individual's age, cohort, and period effect - are confounded (cf. the Panel Study of Income Dynamics). Analytical techniques are necessary to unconfound these three factors and reveal the role of each. As discussed in more detail below, panel data do have some specific problems unconfounding the three major factors. However, for gaining scientific insights into the interplay of processes governing life courses from birth to death, they appear to be a better approach than applying cross-sections. But a mechanical and atheoretical cohort analysis is a useless exercise and statistical innovations alone will not solve the age-period-cohort problem (Blossfeld 1986; Mayer and Huinink 1990).

Most of the above-mentioned difficulties concerning panel studies can be dealt with by sophisticated statistical procedures or more panel waves. However, panel data also lead to a series of deficiencies with respect to the estimation of transition rates (Tuma and Hannan 1984): first, there is the problem of "embeddability," which means that there may be difficulties in embedding a matrix of observed transition probabilities within a continuous-time Markov process (Singer and Spilerman 1976a); second, there is the problem that there may be no unique matrix of transition rates describing the data (Singer and Spilerman 1976b); and third, there is the drawback that the observed matrix of transition probabilities may be very sensitive to sampling and measurement error (Tuma and Hannan 1984). Multiple waves with irregular spacing or shorter intervals between waves can reduce these problems. However, as Hannan and Tuma

(1979) have noted, the more panel and event history data resemble each other, the less problematic modeling becomes.

Lazarsfeld (1948, 1972) was among the first sociologists to propose panel analysis of discrete variables. In particular, he wanted to find a solution to the problem of ambiguity in causation. He suggested that if one wants to know whether a variable X induces change in another variable Y, or whether Y induces change in X, observations of X and Y at two points in time would be necessary. Lazarsfeld applied this method to dichotomous variables whose time-related structure he analyzed in a resulting sixteenfold table. Later on, Goodman (1973) applied log-linear analysis to such tables. For many years, such a cross-lagged panel analysis for qualitative and quantitative variables (Campbell and Stanley 1963; Shingles 1976) was considered to be a powerful quasi-experimental method of making causal inferences. It was also extended to multiwave-multivariable panels to study more complex path models with structural-equation models (Jöreskog and Sörbom 1993). However, it appears that the strength of the panel design for causal inference was hugely exaggerated (Davis 1978). Causal inferences in panel approaches are much more complicated than has been generally realized. Several reasons are responsible for this:

<u>Time until the effect starts to occur.</u> As discussed above, it is important to realize that the role of time in causal explanations does not only lie in specifying a temporal order in which the effect follows the cause in time. It additionally means that a temporal <u>interval is necessary for the cause to have an impact</u> (Kelly and McGrath 1988). The time interval may be short or long, but can never be zero or infinity (Kelly and McGrath 1988). In most of the current sociological theories and interpretations of research findings this interval is left <u>unspecified</u>. In most cases, at least implicitly, researchers assume that the effect takes place almost <u>immediately</u>. Of course, if this is the case, then there seems to be no need for theoretical statements about the time course of causal effects. A single measurement of the effect at some point in time after a cause has been imposed might be sufficient for catching it (see Figure 1a). However, if there is a reason to assume that there is a <u>lag</u> between cause and effect, then a single measurement of the outcome is inadequate for describing the process (see Figure 1b); and the interpretation arrived at on the basis of that single measurement will be a function of the point in time chosen to measure the effect. Thus, a restrictive (implicit) assumption of panel designs is that either <u>cause and effect occur almost simultaneously</u>, or the <u>interval between observations is of approximately the same length as the true causal lag</u>. The greater the discrepancy, the greater the likelihood that the panel analysis will fail to discover the true causal process. Thus, as expressed by Davis (1978), if one does not know the causal lag exactly, panel analysis is not of much utility to establish <u>causal direction</u> or time <u>sequencing of causal effects</u>.

Unfortunately, we rarely, if ever, have enough information about the detailed structure of a social process to specify the true lag precisely.

Temporal shapes of the unfolding effect. While the problem of time-lags is widely recognized in sociology, there is almost no discussion with respect to the temporal shapes of effects (Kelly and McGrath 1988) (see Figure 1). If the effect increases or decreases monotonically (Figure 1c) or linearly, oscillates in cycles (Figure 1.e), or shows any other complicated time-related pattern, then the strength of the observed effect is dependent on the timing of the panel waves. A panel design might be particularly problematic if there are non-monotonic cycles of the effect, because totally opposite conclusions about the effects of the explanatory variable can be arrived at, depending on whether the panel places measurement points at a peak or at an ebb in the curve (see Figures 1d and 1e).

Reciprocal effects with different time-paths. In cases of reciprocal causality, additional problems will arise if the time structure of the effects of X_1 on X_2 and of X_2 on X_1 are different with respect to lags and shapes. In these situations, a panel design might turn out to be completely useless for those wishing to detect such time-related recursive relationships.

Observational data and timing of measurement of explanatory variables. Most sociological research is based on observational data, meaning that manipulation of the timing of the independent variables is generally not possible. For example, if the researcher is going to study the effects of job mobility on marriage behavior, it is impossible to force respondents to change their jobs, say, at the time of the first panel wave. Thus, the longer the interval between panel waves, the more uncertainty there will be regarding the exact point in time when an individual moved to another job and therefore about the point we evaluate in the time-path of the effect (Coleman 1981). The situation may be even more problematic if changes in independent variables are repeatable and several changes are possible between two successive panel waves, as might be the case with job shifts observed in yearly intervals (cf. Sandefur and Tuma 1987). In such panel studies, even the causal order of explanatory and dependent events may become ambiguous.

Observational data and the timing of control variables. Observational studies take place in natural settings and, therefore, offer little control over changes in other important variables and their timing. If these influences change arbitrarily and have time-related effect patterns, then panel studies are useless in disentangling the effect of interest from time-dependent effects of other parallel exogenous processes.

Continuous changes of explanatory and control variables. In observational studies, explanatory and control variables may not only change stepwise from one state to another, but can often change continuously over time. For example, individuals continuously change age, constantly acquire general labor force experience or job-specific human capital if they are employed (Blossfeld and Huinink 1991), are exposed to continuously changing historical conditions (Blossfeld 1986), are steadily changing their social relationships in marital or consensual unions (Blossfeld, De Rose, Hoem, and Rohwer 1993), etc. Even in cases where these continuous changes are not connected with lags or time-related effect patterns, there are deficiencies of panel data concerning their capabilities of detecting time dependence in substantive processes. This is why panel analysis can often not appropriately identify age, period, and cohort effects (Blossfeld 1986).

Therefore, the use of panel data causes an identification problem due to omitted factors whose effects are summarized in a disturbance term. These factors are not stable over time, which means that the disturbance term cannot be uncorrelated with the explanatory variables. Panel analysis therefore critically depends on solutions to the problem of autocorrelation. This problem can be reasonably well tackled by increasing the number of panel waves and modifying their spacing. Panel analysis is particularly sensitive to the length of the time intervals between waves relative to the speed of the process (Coleman 1981). They can be too short, so that too few events will be observed, or too long, so that it is difficult to establish a time-order between events (Sandefur and Tuma 1987). A major advantage of the continuous-time observation design in event history analysis is therefore that it makes the timing between waves irrelevant (Coleman 1968).

Event History Data

For many processes in the social sciences, a continuous measurement of qualitative variables seems to be the only adequate method of assessing empirical change. This is achieved by utilizing an event oriented observation design which records all the changes in qualitative variables and their timing. As shown in Figure 3, the major advantage of event history data is that they provide the most complete data possible on changes in qualitative variables which may occur at any point in time. The observation of events therefore provides an attractive alternative to the observation of states for social scientists.

Event history data, mostly collected retrospectively via life history studies, covers the whole life course of individuals. An example for such a study is the German Life

History Study (GLHS, Mayer and Brückner 1989). Retrospective studies have the advantage of normally being cheaper to collect than panel data. They are also systematically coded to one framework of codes and meanings (Dex 1991). But retrospective (in contrast to prospective) studies suffer from several limitations that have been increasingly acknowledged (Medical Research Council 1992):

Non-factual data. It is well-known that retrospective questions concerning motivational, attitudinal, cognitive, or affective states are particularly problematic because the respondents can hardly recall the timing of changes in these states accurately (Hannan and Tuma 1979). This type of data is not verifiable even in principle, as these states exist only in the minds of the respondents and are only directly accessible, if at all, to the respondent concerned (Sudman and Bradburn 1986). For these non-factual data, panel studies have the advantage of being able to repeatedly record current states of the same individual over time. Thus, for studies aiming to model the relationship between attitudes and behavior over time, panel observations of attitudinal states, combined with retrospective information on behavioral events since the last sweep, appear to be an appropriate design.

Recall problems with regard to behavior or facts. Behavioral or factual questions ask the respondents about characteristics, things they have done, or things that have happened to them which in principle are verifiable by an external observer. Most surveys (cross-sectional, panel or event oriented) elicit retrospective information on behavior or facts (e.g. by asking people about their education, social origin etc.), so that the disadvantages of retrospection are only a matter of degree. However, event history studies are particularly ambitious (see Mayer and Brückner 1989). They try to collect continuous records of qualitative variables that have a high potential for bias because of their strong reliance on memory. However, research on the accuracy of retrospective data shows that individuals' marital and fertility histories, family characteristics and education, health service usage, and employment history can be collected to a reasonable degree of accuracy. A very good overview concerning the kinds of data which can be retrospectively collected, the factors affecting recall accuracy, and the methods improving recall has been presented by Shirley Dex (1991).

Unknown factors. Retrospective designs cannot be used to study factors involving variables that are not known to the respondent (e.g. emotional and behavioral problems when the respondent was a child). In such cases, panel studies are indispensable (Medical Research Council 1992).

Limited capacity. There is a limit to respondents' tolerance for the amount of data which can be collected on one occasion (Medical Research Council 1992). A carefully corsetted panel design can therefore provide a broader coverage of variables (if these are not unduly influenced by variations at the time of assessment).

Only survivors. Due to their nature, retrospective studies must be based on survivors. Thus, those subjects who have died or migrated from the geographical area under study will necessarily be omitted. If either is related to the process (as often may be the case), biases will arise. This problem is particularly important for retrospective studies involving a broad range of birth cohorts, like the GLHS or international migration studies (Blossfeld 1987).

Misrepresentation of specific populations. Retrospective studies also systematically misrepresent specific populations. For example, Duncan (1966) has shown that if men are asked about their fathers, men from earlier generations who had no sons, or whose sons died or emigrated are not represented in a retrospective father-son mobility table.

4. Conclusion

The aim of this paper has been to show that - compared to alternative observation plans - the collection of event history data is an extremely useful approach for uncovering causal relationships or mapping out systems of causal relations. Event history models provide a natural basis for a causal understanding of social processes because they relate the change in future outcomes to conditions in the past at each point in time and enable the researcher to predict future changes on the basis of past observations at each moment of the process (Aalen 1987).

In the past, event history data normally have been collected retrospectively leading to specific problems described above. To avoid these problems (or to diminsh their relevance), a mixed design employing a follow-up (or "catch-up") and a follow-back strategy appears to combine the strengths of panel designs with the virtues of retrospective studies. Therefore, in modern panel studies event histories are collected retrospectively for the period before the panel started and between the successive panel waves. Sometimes, complete administrative records also contain time-related information about events in the past. In any case, with regard to cross-sectional and tranditional panel data, all of these procedures (retrospective, combined follow-up and back-up, or available registers) offer a comparatively superior opportunity for modeling social processes, regardless of which method is selected.

References

Aalen, O. O. (1987): "Dynamic modelling and causality," in: Scand. Actuarial Journal, 12:177-190.

Arminger, G., C.C. Clogg, and M.E. Sobel (eds.) (1995): Handbook of statistical modeling for the social and behavioral sciences, Plenum.

Arminger, G. and F. Müller (1990): Lineare Modelle zur Analyse von Paneldaten, Opladen.

Barron, D. N. (1993): "The analysis of count data: overdispersion and autocorrelation," in: Sociological Methodology, pp. 179-220.

Blossfeld, H.-P. (1986): "Career opportunities in the Federal Republic of Germany: a dynamic approach to the study of life-course, cohort, and period effects," in: European Sociological Review, 2:208-225.

Blossfeld, H.-P. (1987): "Labor market entry and the sexual segregation of careers in the FRG," in: American Journal of Sociology, 93:83-118.

Blossfeld, H.-P. (1989): Kohortendifferenzierung und Karriereprozeß. Eine Längsschnittstudie über die Veränderung der Bildungs- und Brerufschancen im Lebenslauf, Frankfurt a. M. and New York.

Blossfeld, H.-P. (ed.) (1995a): Family formation in modern societies and the new role of women, Boulder et al. (in print).

Blossfeld, H.-P. (ed.) (1995b): Between equalization and marginalization. Part-time working women in Europe and the United States of America, Oxford (forthcoming).

Blossfeld, H.-P., A. DeRose, J. Hoem, and G. Rohwer (1993): Education, modernization, and the risk of marriage disruption: differences in the effect of women's educational attainment in Sweden, West-Germany, and Italy," Stockholm Research Reports in Demography, no. 76, Stockholm (To appear in: K. Oppenheim Mason and An-Magritt Jensen (eds.): Gender and family change in industrialized countries, Oxford).

Blossfeld, H.-P., A. Hamerle and K. U. Mayer (1989): Event history analysis, Hillsdale (N.J.).

Blossfeld, H.-P. and J. Huinink (1991): "Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation," in: American Journal of Sociology, 97:143-168.

Blossfeld, H.-P., D. Manting, and G. Rohwer (1993): "Patterns of change in family formation in the Federal Republic of Germany and the Netherlands: some consequences for solidarity between generations", in: H. A. Becker and P. L. J. Hermkens (eds.): Solidarity of generations. Demographic, economic and social change, and its consequences, Amsterdam. pp. 175-196.

Blossfeld, H.-P. and G. Rohwer (1994): "West-Germany," in: H.-P. Blossfeld (ed.): Between equalization and marginalization. Part-time working women in Europe and the United States of America, Oxford (forthcoming).

Blossfeld, H.-P. and G. Rohwer (1995): Techniques of event history modeling. New approaches to causal analysis, Hillsdale (NJ) (in print).

Bollen, K. A. (1989): Structural equations with latent variables, New York.

Campbell, D. T. and J. C. Stanley (1963): Experimental and quasi-experimental designs for research, Chicago.

Campbell, R. T., E. Mutran, and R. Nash Parker (1987): "Longitudinal design and longitudinal analysis. A comparison of three approaches," in: Research on Aging, 8:480-504.

Chamberlain, G. (1984): "Panel data," in. Z. Griliches and M. D. Intriligator (eds.): Handbook of econometrics, vol. II, pp. 1247-1317. Amsterdam.

Clogg, C. C. and G. Arminger (1993): "On strategy for methodological analysis", in: Sociological Methodology, 23:57-74.

Coleman, J. S. (1968): "The mathematical study of change," in: Hubert, M. and A. B. Blalock (eds.): Methodology in Social Research, New York, pp. 428-478.

Coleman, J. S. (1981): Longitudinal data analysis, New York.

Coleman, J. S. (1990): Foundations of social theory, Cambridge.

Coleman, J.S. and L. Hao (1989): "Linear systems analysis: macrolevel analysis with microlevel data," in: Sociological Methodology, 19:395-422.

Davies, R. B. (1987): "The limitations of cross-sectional analysis," in: Crouchley, R. (ed.): Longitudinal data analysis, Aldershot et al. pp. 1-15.

Davies, R. B. and R. Crouchley (1985): "Control for omitted variables in the analysis of panel and other longitudinal data," in: Geographical Analysis, 17:1-15.

Davis, J. A. (1978): "Studying categorical data over time," in: Social Science Research, 7:151-179.

Dex, S. (ed.)(1991): Life and work history analyses: Qualitative and quantitative developments, London and New York.

DiPrete, T.A. (1993): "Industrial restructuring and the mobility response," in: American Sociological Review, 58:74-96

Duncan, O.D. (1966): "Methodological issues in the analysis of social mobility," in: N. Smelser and S.M. Lipset (eds.): Social structure and social mobility, Chicago.

Eells, E. (1991): Probabilistic causality, Cambridge.

Engel, U. and J. Reinecke (1994): Panelanalyse, Berlin.

Faulbaum, F. and P. M. Bentler (1994): "Causal modeling: some trends and perspectives," in. I. Borg and P.O. Mohler (eds.): Trends and perspectives in empirical social research, pp. 224-249, Berlin.

Flinn, Ch. J. and J. J. Heckman (1982): "Models for the analysis of labor force dynamics," in: R. Basmann and G. Rhodes (eds.): Advances in econometrics, Greenwich (C.T.), pp. 35-95.

Freeman, R. A. (1991): "Statistical analysis and shoe leather," in: Sociological Methodology, 21:291-313.

Glenn, N.D. (1977): Cohort analysis, Beverly Hills (C.A.).

Goldthorpe, J. H. (1991): "The use of history in sociology: reflections on some recent tendencies," in: Brithish Journal of Sociology, 42:211-230.

Goldthorpe, J. H. (1994): The quantitative analysis of large-scale data-sets and rational action theory: For a sociological alliance," in: Oxford, Nuffield College (mimeo).

Goodman, L.A. (1973): "Causal analysis of data from panel studies and other kind of surveys," in: American Journal of Sociology, 78:1135-1191.

Hannan, M. T. and N.B. Tuma (1979): "Methods for temporal analysis," in: Ann. Rev. Sociol., 5:303-328.

Heckman, J.J. and R.J. Willis (1977): "A beta-logistic model for the analysis of sequential labor force participation by married women," in: Journal of Political Economy, 85:27-58.

Heinz, W. R. (ed.) (1991a): Theoretical advances in life course research. Status passages and the life course, vol. I, Weinheim.

Heinz, W. R. (ed.) (1991b): The life course and social change. Status passages and the life course, vol. II, Weinheim.

Heinz, W. R. (ed.) (1992): Institutions and gate keeping in the life course. Status passages and the life course, vol. III, Weinheim.

Holland, P. W. (1986): "Statistics and causal inference," in: Journal of the American Statistical Association, 81:945-960.

Hsiao, C. (1986): Analysis of panel data. Cambridge.

Huinink, J. (1989): Mehrebenenanalyse in den Sozialwissenschaften, Wiesbaden.

Hunt, M. (1985): Profiles of social research. The scientific study of human interactions, New York.

Jöreskog, K. G. and D. Sörbom (1993): LISREL 8. Structural equation modeling with the SIMPLIS command language, Chicago.

Kelly, J. R. and J. E. McGrath (1988): On time and method, Newbury Park et al.
Lazarsfeld, P.F. (1948): "The use of panels in social research," in: Proceedings of the American Philosophy of Sociology, 92:405-410.

Lazarsfeld, P.F. (1972): "Mutual relations over time of two attributes: a review and integration of various approaches," in: M. Hammer, K. Salzinger, and S. Sutton (eds.): Psychopathology, New York.

Lieberson, S. (1985): Making it count. The improvement of social research and theory. Berkeley et al.

Lieberson, S. (1991): "Small n's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases," in: Social Forces, 70:307-320.

Magnusson, D. and L. R. Bergmann (eds.) (1990): Data quality in longitudinal research, Cambridge.

Magnusson, D., L. R. Bergmann, and B. Törestad (eds.) (1991): Problems and methods in longitudinal research: stability and change, Cambridge.

Manting, D. (1994): Dynamics in marriage and cohabitation. An inter-temporal, life course analysis of first union formation and dissolution, Amsterdam.

Marini, M. M. and B. Singer (1988): "Causality in the social sciences," in: C. C. Clogg (ed.): Sociological Methodology, pp. 347-409.

Mason, W.M. and S.E. Fienberg (1985) (eds.): Cohort analysis in social research.

Mayer, K.U. (ed.)(1990): Lebensverläufe und sozialer Wandel, Sonderheft 31, Kölner Zeitschrift für Soziologie und Sozialpsychologie, Opladen.

Mayer, K. U. (1991): "Life courses in the welfare state," in: W. R. Heinz (ed.): Theoretical advances in life course research, vol. I. pp. 171-186, Weinheim.

Mayer, K. U. and E. Brückner (1989): Lebensverläufe und Wohlfahrtsentwicklung. Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1929-31, 1939-41, 1949-51, Materialien aus der Bildungsforschung, no. 35, Max-Planck-Institut für Bildungsforschung, Berlin.

Mayer, K. U. and J. Huinink (1990): "Age, period, and cohort in the study of the life course: a comparison of classical A-P-C-analysis with event history analysis or farewell to Lexis?," in. D. Magnusason and L. R. Bergmann (eds.): Data quality in longitudinal research, pp. 211-232, Cambridge.

Mayer, K. U. and N. B. Tuma (eds.) (1990): Event history analysis in life course research, Madison (W.I.)

Medical Research Council (1992): Review of longitudinal studies, London (mimeo).

Murphy, M. (1991): The family life cycle, London School of Economics, London (mimeo).

Pickles, A. R. and R. B. Davies (1989): "Inference from cross-sectional and longitudinal data for dynamic behavioural processes," in: J. Hauer et al. (eds.): Urban dynamics and spatial choice behaviour, pp. 81-104.

Rodgers, W.L. (1982): "Estimable functions of age, period, and cohort effects,", in: American Sociological Review, 47:774-787.

Rosenthal, Leslie (1991): "Unemployment incidence following redundancy; the value of longitudinal approaches," in: Dex, S. (ed.): Life and work history analyses: Qualitative and quantitative developments, London and New York, pp. 187-213

Sandefur, G. D. and N. B. Tuma (1987): "How data type affects conclusions about individual mobility," in: Social Science Research, 16:301-328.

Shavit, Y. and H.-P. Blossfeld (eds.): Persistent inequality. Changing educational attainment in thirteen countries, Boulder et al.

Shingles, R. D. (1976): "Causal inference in cross-lagged panel analysis," in: Political Methodology, 3:95-133.

Singer, B. and S. Spilerman (1976a): "The representation of social processes by Markov models," in: American Journal of Sociology, 82:1-54.

Singer, B. and S. Spilerman (1976b): "Some methodological issues in the analysis of longitudinal surveys," in: Annals of Economic and Social Measurement, 5:447-474.

Sudman, S. and N. M. Bradburn (1986): Asking questions, San Francisco et al.

Tuma, N. B. and M. T. Hannan (1984): Social dynamics. Models and Methods. Orlando et al.

Yamaguchi, K. (1991): Event history analysis, Newbury Park (C.A.) et al.

Sonderforschungsbereich 186
der Universität Bremen

Statuspassagen und Risikolagen
im Lebensverlauf

Causal Inference, Time and Observation Plans in the Social Sciences

von

Hans-Peter Blossfeld
und
Götz Rohwer

Arbeitspapier Nr. 36

Bremen 1996

Preface

The B6 project within the Special Collaborative Program on "Status Passages and Risks in the Life Course" uses advanced quantitative longitudinal methodolgies. A great deal of work in this project is directed toward the development of new techniques and methods.

This paper demonstrates that the opportunity for assessing causal inferences varies strongly with the type of observation available to the social scientist. The data structure (which can be cross-sectional, panel or event oriented) determines the extent to which the researcher is forced to make untested assumptions in the process of establishing relevant empirical evidence that can serve as a link in a chain of reasoning about causal mechanisms.

The authors of this paper stress that the collection of event history data offer a comparatively superior approach for uncovering causal relationships or mapping out systems of causal relations. This is because event history models relate the change in future outcomes to conditions in the past at each point in time.

Prof. Dr. Ansgar Weymann
Chai, Special Collaborative Programme No. 186

Table of Contents

1. Introduction

The investigation of causal relationships is an important but difficult scientific endeavor. The opportunity for assessing causal inferences varies strongly with the type of observation available to the social scientist. This is because the data structure determines the extent to which the researcher is forced to make untested assumptions when he or she is trying to establish relevant empirical evidence that can serve as a link in a chain of reasoning about causal mechanisms (Goldthorpe 1994).

In this paper, we first discuss the role of time in causal inferences in the social sciences. This will help us to recast a number of research problems so that fundamental design issues can be addressed more directly. Then, based on various examples, we will describe in detail how different observation plans (cross-secional, panel, and event history oriented) affect causal analysis. Although longitudinal data are no panacea, they are obviously more effective in causal analysis and have less inferential limitations.

We will limit our attention to data generated by continuous-time, discrete-state substantive processes. These processes have been characterized by James Coleman (1981:6) in the following general way: (1) there are units - which may be individuals, organizations, societies, or whatever - that change from one discrete state to another; (2) these changes (or events) can occur at any point in time and are not restricted to predetermined points in time; and (3) there are time-constant and/or time-dependent factors influencing the events. Examples of such process type can be given for a wide variety of social research fields (see, e.g., Blossfeld, Hamerle, and Mayer 1989; Blossfeld and Rohwer 1995). Consider, for example, a person's job career. The job history may be described as the time spent in the first job and the date the person entered into this spell, the type and duration in the following job or the kind of non-employment and its duration, and so on.

2. Causal Statements in the Social Sciences

In this section, we will focus on the general role of time in causal inferences and also show that the idea of a causal effect can be represented as a change in the transition rate, if the dependent variable is discrete and can change its state at any time.

Correlation and Causation

To begin with, statements about causation should be distinguished from statements about association. In making correlational inferences, one can be satisfied to observe how the values of one variable are associated with the values of other variables over the population under study and perhaps over time. In this context, time is only important insofar as it determines the population under analysis or specifies the operational meaning of a particular variable (Holland 1986). Statements about associations describe what has happened. They are quite different from causal statements and are designed to give information about how events are produced or conditioned by other events.

Sometimes social scientists argue that because the units of sociological analysis continuously learn and change and involve actors with goals and beliefs, sociology can at best only provide systematic descriptions of phenomena at various points in history. This position is based on the view that causal statements about events are only possible if they are regulated by "eternal," time-less laws (Kelly and McGrath 1988). Of course, the assumption that such laws can be established with regard to social processes can reasonably be disputed. However, we are not forced to accept a simple contrast: either describing contingent events or assuming "eternal" laws. Many social phenomena show <u>systematic temporal variations and patterned regularities</u> under specific conditions that themselves are a legitimate focus of our efforts to understand social change (Kelly and McGrath 1988). Thus, sociology can do more than just describe the social world. This paper therefore emphasizes the usefulness of techniques of event history modeling as "new" approaches to the investigation of causal explanations.[1]

Causal Mechanisms and Substantive Theory

The identification of causal mechanisms has been one of the classic concerns in sociology. Causal statements are made to explain the occurrence of events, to understand why particular events happen, and to make predictions when the situation changes (Marini and Singer 1988). Although sociologists sometimes seem to be

---

[1] We speak of a "new" approach just to emphasize the contrast to traditional "causal analysis" based on structural equation models which are basically time-less models. See the discussion in Bollen (1989), Campbell, Mutran and Parker (1987), or Faulbaum and Bentler (1994).

opposed to using the word "cause," they are far less reluctant to apply very similar words like "force," "agency," or "control," when trying to understand social phenomena.

There is consensus in the fact that causal inferences cannot simply and directly be made from empirical data, regardless of whether they are collected through ingenious research designs or summarized by particularly advanced statistical models. Thus, using event history observation plans and event history analysis models per se will not allow us to prove causality, as is the case for all other statistical techniques. However, as we will see in the next section, event-oriented observation designs offer richer information and, as we will try to demonstrate in this paper, event history models provide more appropriate techniques for exploring causal relations.

It seems useful to treat causality as being a property of theoretical statements rather than the empirical world itself (Goldthorpe 1994). In sociology, causal statements are based primarily on substantive hypotheses which the researcher develops about the social world. In this sense, causal inference is theoretically driven (Freedman 1991) and it will always reflect the changing state of sociological knowledge in a field.[2] Of course, descriptive statements are also dependent on theoretical views guiding the selection processes and providing the categories underlying every description. The crucial point is however that causal statements need a theoretical argument specifying the particular mechanism of how a cause produces an effect or, more generally, in which way interdependent forces affect each other in a given setting over time.

Therefore, the important task of event history modeling is not to demonstrate causal processes directly, but to establish relevant empirical evidence that can serve as a link in a chain of reasoning about causal mechanisms (Goldthorpe 1994). In this respect, event history models might be particularly helpful instruments because they allow a time-related empirical representation of the structure of causal arguments.

Attributes, Causes and Time-Constant Variables

Holland (1986) tried to establish some links between causal inference and statistical modeling. In particular, he emphasized that for a conception of causality it is essential that each unit of a population must be exposable to any of the various levels of a cause, at least hypothetically. He argues, for example, that the schooling a student

---

[2] Causal relations are always identified against the background of some field, and specification of a field is critical to the identification of an observed relation (Marini and Singer 1988).

receives can be a cause of the student's performance on a test, whereas the student's race or sex cannot. In the former case it seems possible to contemplate measuring the causal effect, whereas in the latter cases, where we have the enduring attributes of a student, all that can be discussed is association (see also Yamaguchi 1991).

We agree with Holland that it is essential for causal statements to imply counterfactual reasoning: if the cause had been different, there would have been another outcome, at least with a certain probability. In this sense, counterfactual statements reflect imagined situations. It is not always clear, however, which characteristics of a situation can sensibly be assumed to be variable, i.e. can be used in counterfactual reasoning, and which characteristics should be regarded as fixed. At least to some degree, the distinction depends on the field of investigation. For example, from a sociological point of view what is important with regard to sex is not the biological attributes per se, but the social meaning attached to these attributes. The social meaning of these attributes can change regardless of whether their biological basis changes or not. For example, societal rules might change to create more equality between the races or sexes. We therefore think that in sociological applications counterfactuals can also be meaningfully applied to such attributes. They can be represented as time-constant "variables" in statistical models to investigate their possible impact on some outcome to be explained. It is, however, important to be quite explicit about the sociological meaning of causal statements which involve references to biological or ethnic attributes. There is, for example, no eternal law connecting gender and/or race with wage differentials. But probably there are social mechanisms which connect gender and ethnic differences with different opportunities in the labor market.

Causes and Time-Dependent Variables

The meaning of the counterfactual reasoning of causal statements is that causes are states which could be different from what they actually are. However, the consequences of conditions that could be different from their actual state are obviously not observable.[3] It means that it is simply impossible to observe the effect that would have happened on the same unit of analysis, if it were exposed to another condition at the same time. To find an empirical approach to causal statements, the researcher must look at conditions which actually do change in time. These changes are events. More formally, an event is a change in a variable, and this change must happen at a specific point in time. This implies that the most obvious empirical representation of causes is

[3] Holland (1986) calls this "the fundamental problem of causal inference."

in terms of variables that can change their states over time. This statement is linked very naturally with the concept of time-dependent covariates in event history analysis. The role of a time-dependent covariate in this approach is to indicate that a (qualitative or metric) causal factor has changed its state at a specific time and that the unit under study is exposed to another causal condition. For example, in the case of gender the causal events might be the steps in the acquisition of gender roles over the life course or the exposure to sex-specific opportunities in the labor market at a specific historical time. Thus, a time-constant variable "gender" should ideally be replaced in an empirical analysis by time-changing events assumed to produce sex-specific differences in the life history of men and women. Of course, in empirical research that is not always possible, so that one very often has to rely on time-constant "variables" as well. However, it is important to recognize that for these variables the implied longitudinal causal relation is not examined. For example, if we observe an association amongst people with different levels of educational attainment and their job opportunities, then we can normally draw the conclusion that changes in job opportunities are a result of changes in educational attainment level. The implied idea is the following: if we started having people with the lowest educational attainment level and followed them over the life course, they would presumably differ in their rates to attaining higher levels of educational attainment and this would produce changes in job opportunities. Whether this would be the case for each individual is not very clear from a study that is based on people with different levels of educational attainment. In particular, one would expect that the causal relationship between education and job opportunities would radically be altered if all people acquired a higher (or the highest) level of educational attainment.[4] Thus, the two statements - the first about associations across different members of a population and the second about dependencies in the life course for each individual member of the population - are quite different; one type of statement can be empirically true while the other one can be empirically false. Therefore, statements of the first type cannot be regarded as substitutes for statements of the second type. However, since all causal propositions have consequences for longitudinal change (see Lieberson 1985), only time-changing variables provide the most convincing empirical evidence of causal relations.[5]

---

4 A longitudinal approach would provide, however, the opportunity to study these kinds of changes in the causal relationships over time.

5 There is also another aspect that is important here (see Lieberson 1985): causal relationships can be symmetric or asymmetric. In examining the causal influence of a change in a variable X on a change in a dependent variable Y, one has to consider whether shifts to a given value of X from either direction have the same consequences for Y. For example, rarely do researchers consider whether an upward shift on the prestige scale, say from 20 to 40, will lead to a different

Time Order and Causal Effects

We can summarize our view of causal statements in the following way:

$$\Delta X_t \rightarrow \Delta Y_{t'} \qquad t < t'$$

meaning that a change in variable X_t at time t is a cause of a change in variable $Y_{t'}$ at a later point in time, t'. It is not implied, of course, that X_t is the only cause which might affect $Y_{t'}$. So we should speak of causal conditions to stress that there might be, and normally is, a quite complex set of causes.[6] Thus, if causal statements are studied empirically, they must intrinsically be related to time. There are three important aspects. First, to speak of a change in variables necessarily implies reference to a time axis. We need at least two points in time to observe that a variable has changed its value. Of course, at least approximately, we can say that a variable has changed its value at a specific point in time.[7] Therefore, we use the symbols to refer to changes in the values of the time-dependent variable ΔX_t and the state variable ΔY_t at time t. This leads to the important point that causal statements relate changes in two (or more) variables.

---

outcome of Y (say family decisions) than would a downward shift of X from 60 to 40. In other words, most researchers assume symmetry. However, even if a change is reversible, the causal process may not be. The question is: if a change in a variable X causes a change in another one, Y, what happens to Y if X returns to its earlier level? "Assuming everything else is constant, a process is reversible, if the level of Y also returns to its initial condition; a process is irreversible if Y does not return to its earlier level. Observe that it is the process - not the event - that is being described as reversible or irreversible." (Lieberson 1985:66)

6 It is important to note here that the effect of a variable X is always measured relative to other causes. A conjunctive plurality of causes occurs if various factors must be jointly present to produce an effect. Disjunctive plurality of causes, on the other hand, occurs if the effect is produced by each of several factors alone, and the joint occurrence of two or more factors does not alter the effect (see the extensive discussion in Marini and Singer (1988)).

7 Statements like this implicitly refer to some specification of "point in time." The meaning normally depends on the kind of events which are to be described, for instance, a marriage, the birth of a child, or to become unemployed. In event history text books, normally a continuous time axis for purposes of mathematical modeling is assumed (see Blossfeld and Rowher 1995). This should however be understood as an idealized way of representing social time. Here we are using mathematical concepts to speak about social reality, so we will disregard the dispute about whether time is "continuous" (in the mathematical sense of this word), or not.

Second, there is a time ordering between causes and effects. The cause must precede the effect in time: t < t', in the formal representation given above. This seems to be generally accepted.[8] As an implication, there must be a temporal interval between the change in a variable representing a cause, and a change in the variable representing a corresponding effect. It is important to realize that the role of time in causal explanations does not only lie in specifying a temporal order in which the effect follows the cause in time. It additionally implies that a temporal interval is necessary for the cause to have an impact (Kelly and McGrath 1988). In other words, if the cause has to precede the effect in time, it takes some finite amount of time for the cause to produce the effect. The time interval may be very short or very long, but can never be zero or infinity (Kelly and McGrath 1988). Some effects take place almost instantaneously. For example, if the effect occurs at microsecond intervals, then the process must be observed in these small time units to uncover causal relations. However, some effects may occur in a time interval too small to be measured by any given methods, so that cause and effect seem to occurat the same point in time. Apparent simultaneity is often the case in those social science applications where basic observation intervals are relatively crude (e.g. days, months, or even years), such as, for example, yearly data about first marriage and first childbirth (Blossfeld, Manting and Rohwer 1993). For these parallel processes, the events "first marriage" and "first childbirth" may be functionally interdependent, but whether these two events are observed simultaneously or successively depends on the degree of temporal refinement of the scale used in making the observations. Other effects need a long time until they start to occur. Thus, there is a delay or lag between cause and effect that must be specified in an appropriate causal analysis. However, in most of the current sociological theories and interpretations of research findings this interval is left unspecified.

This immediately leads to a third point. In addition to the question of how long the delay between the timing of the cause and the beginning of the unfolding of the effect is, there might be different shapes of how the causal effect Y_t unfolds over time. While the problem of time-lags is widely recognized in social science literature, there is almost no information with respect to the temporal shapes of effects (Kelly and McGrath 1988). Social scientists seem to be quite ignorant with respect to the fact that causal effects could be highly time-dependent, too. The panels of Figure 1 illustrate several possible shapes these effects may trace over time. In Figure 1a, there is an almost all-at-once change that is then maintained; in Figure 1b, the effect occurs with some lengthy time-lag and is then time-invariant; in Figure 1c, the effect starts almost

---

[8] See, for instance, the discussion in Eells (1991, Ch.,5).

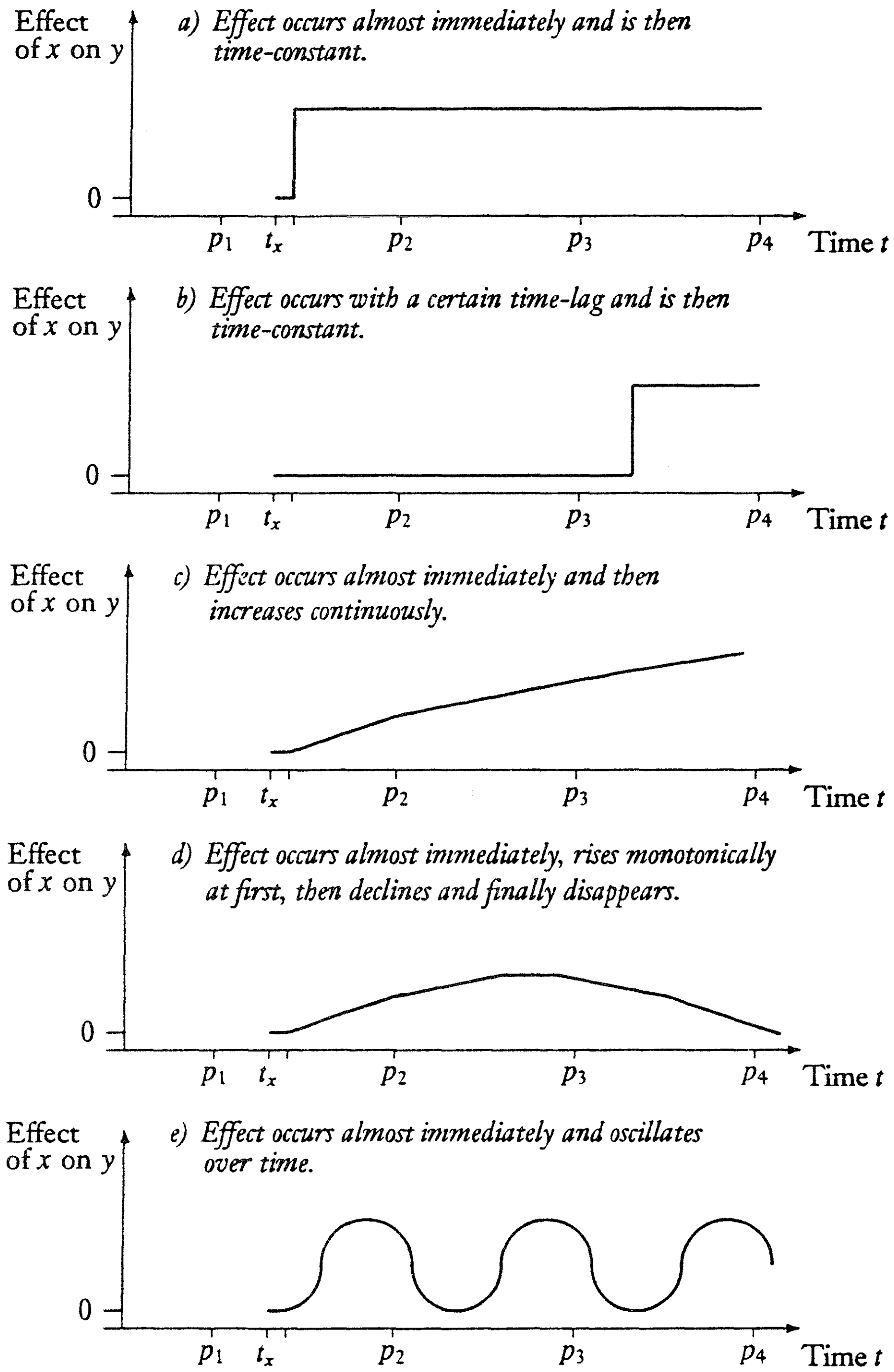

Figure 1 Different temporal shapes of how a change in a variable x, occurring at point in time t_x, effects a change in a variable y.

immediately and then gradually increases; in Figure 1d, there is an almost all-at-once increase, which reaches a maximum after some time and then decreases; finally, in Figure 1e, a cyclical effect pattern over time is described. Thus, an appropriate understanding of causal relations between variables should take into account that the causal relationship itself may change over time. This seems particularly important in sociological applications of causal reasoning. In these applications we generally cannot rely on the assumption of eternal, time-less laws but have to recognize that the causal mechanisms may change during the development of social processes.

Actors and Probabilistic Causal Relations

It seems agreed that social phenomena are always directly or indirectly based on actions of individuals. This clearly separates the social from the natural sciences. Sociology therefore does not deal with associations among variables per se, but with variables that are associated via acting people. There are at least three consequences for causal relations: First, in methodological terms, this means that if individuals relate causes and effects through their actions, then research on social processes should at best be based on individual longitudinal data (Coleman and Hao 1989; Coleman 1990). This is why life history data on individuals, and not aggregated longitudinal data, provide the most appropriate information for the analyses of social processes. Only with these data can one trace the courses of action at the level of each individual over time. Second, in theoretical terms it means that the explaining or understanding of social processes requires a time-related specification of (1) the past and present conditions under which people act,[9] (2) the many and possibly conflicting goals that they pursue at the present time, (3) the beliefs and expectations guiding the behavior, and (4) the actions that probably will follow in the future.[10]

---

[9] These conditions are, of course, heavily molded by social structural regularities in the past and the present. Sociology must always be a historical discipline (Goldthorpe 1991).

[10] Sometimes it is argued that, since human actors act intentionally and behavior is goal-oriented, the intentions or motives of actors to bring about some effect in the future causes the actor to behave in a specific way in the present (Marini and Singer 1988). This does not however contradict a causal view. One simply has to distinguish intentions, motives or plans as they occur in the present from their impact on the behavior which follows their formation temporally, and from the final result, as an outcome of the behavior. An expectation about a future state of affairs should clearly be distinguished from what eventually happens in the future. Therefore, the fact that social agents can behave intentionally, based on expectations, does not reverse the time order underlying our causal statements.

Third, if it is people that are doing the acting, then causal inference must also take into account the free will of individuals. This introduces an essential element of indeterminacy into causal inferences. This means that in sociology we can only reasonably account for and model the generality but not the determinacy of behavior. The aim of substantive and statistical models must therefore be to capture common elements in the behavior of people, or patterns of action that recur in many cases (Goldthorpe 1994). This means that in sociological applications randomness has to enter as a defining characteristic of causal models. We can only hope to make sensible causal statements about how a given or (hypothesized) change in variable X_t in the past affects the probability of a change in variable $Y_{t'}$ in the future. Correspondingly, the basic causal relation becomes

$$\Delta X_t \rightarrow \Delta \Pr(\Delta Y_{t'}) \qquad t < t' \qquad (1.1)$$

This means that a change in the time-dependent covariate X_t will change the probability that the dependent variable $Y_{t'}$ will change in the future ($t' > t$). In sociology, this interpretation seems more appropriate than the traditional deterministic approach. The essential difference is not that our knowledge about causes is insufficient because it only allows probabilistic statements, but that the causal effect to be explained is a probability. Thus, probability in this context is not just a technical term anymore, but is considered as a theoretical one: it is the propensity of social agents to change their behavior.

Causal Statements and Limited Empirical Observations

A quite different type of randomness related to making inferences occurs if causal statements are applied to real-world situations in the social sciences. There are at least four additional reasons to expect further randomness in empirical studies. These are basically the same ones that occur in deterministic approaches and are well-known from traditional regression modeling (Lieberson 1991). The first one is measurement error, a serious problem in empirical social research, which means that the observed data deviate somewhat from the predicted pattern without invalidating the causal proposition. The second reason is particularly important in the case of non-experimental data. It is often the case that complex multivariate causal relations operate in the social world. Thus, a given outcome can occur because of the presence of more than one influencing factor. Moreover, it may also not occur at times because the impact of one independent variable is outweighed by other influences working in the opposite direction. In these situations, the observed influence of the cause is only approximate, unless one can control for the other important factors. The third motive is that sociologists often do not know or are not able to measure all of the important factors. Thus, social scientists have to relinquish the idea of a complete measurement of causal effects, even if they would like to make a deterministic proposition. Finally, sometimes chance affects observed outcomes in the social world. It is not important here to decide whether chance per se exists or whether it is only a surrogate for the poor state of our knowledge of additional influences and/or inadequate measurement.

In summary, these problems imply that social scientists can only hope to make empirical statements with a probabilistic character. This situation normally leads to identification problems. Without strong assumptions about missing information and errors in the available data, it is generally not possible to find definite statements about causal relations (see Blossfeld and Rohwer 1995).

A Simplistic Conception of Causal Relations

At this point it is important to stress that the concept of causal relation is a rather special abstraction implying a time-related structure that does not immediately follow from our sensory impressions. Consider the following example in Figure 2 where we characterize the necessary time-related observations of a unit being affected by a causal effect. This figure shows that an empirical representation of the most simple causal effect, (i.e. (1) where the condition X_t changes (from one state $X_{t1}= x_1$ to another one $X_{t2} = x_2$) and (2) is then constant afterwards, (3) the change in Y_t (from $Y_{t2} = y_1$ to $Y_{t3} = y_2$) takes place almost instantaneously and (4) is then also time-constant afterwards) needs at least three points in time where the researcher must note the states of the independent and dependent variables, respectively.[11] This is because, if we assume that a change in the independent variable X_t has taken place at t_2, then, to be able to fix the particular change in the condition in the past, we need to know the state of the independent variable X_t at an earlier time, t_1 (see Figure 2). For the dependent variable Y_t we need an observation before the effect has started to occur. Assuming everything else is constant, this observation can be made, at the latest at point t_2 because the effect has to follow the cause in time. To evaluate whether the hypothesized effect has indeed taken place at a later time, t_3, we must again note the state of the dependent variable Y_t. Thus, a simplistic representation of a causal effect exists when we compare the change in the observations for the independent variable in the past and the present with the change in the observations

[11] This example is instructive because Lazarsfeld (1948,1972), and many others after him have argued that for causal inferences two observations of the units would be sufficient.

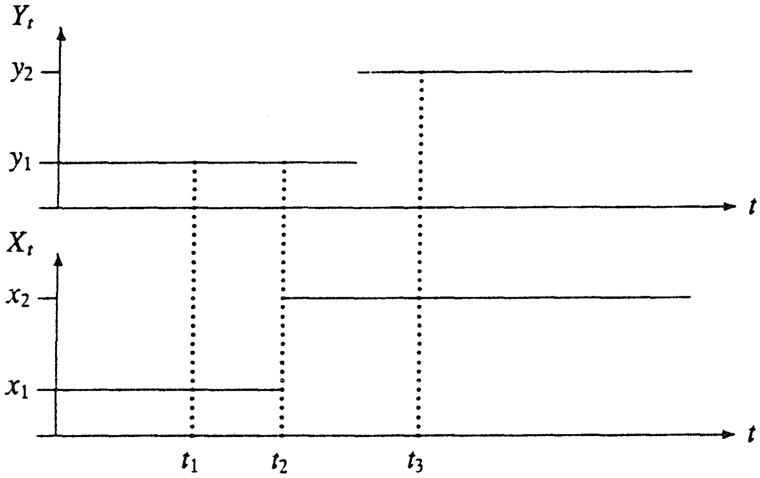

Figure 2 Observation of a simple causal effect.

for the dependent variable in the present and in the future and link both changes with a substantive argument.[12]

However, this is only a simple and fairly unrealistic example of a causal effect. In the case of observational data in the social sciences, where there are many (qualitative and metric) causal variables that might change their values at any point in time, when their causal effects might have various delays and different shapes in time (see Figure 1), then the quantity of the observed causal effect as shown in Figure 1 will strongly depend on when the measurements at the three points in time are taken.

Thus, what can we say about the causal effect(s) at any given point in time if the situation is more complex? A paradox occurs: the concept of causal effect depends intrinsically on comparisons between changes in both the independent and dependent variables in at least three points in time. Yet the concept of causal effect should itself reflect a state of a unit of observation at any point in time as being an appropriate one in real empirical situations. Thus, what is still needed in our discussion is a concept that represents the quantity of the causal effect at any point in time.

Causal Effects and Changes in Transition Rates

If the dependent variable is discrete and can change its state at any time, then the transition rate framework offers a time-point-related representation for the causal effect. We briefly want to develop this idea.

[12] Indeed such a simplistic idea of the causal effect is the basis of all panel designs, as shown in the next section.

Let us first start with the dependent variable, Y_t, and its changes in the future (as a consequence of a change in a causal factor). In particular, we are interested in changes of states occupied by the units of analysis. The state space is assumed to be discrete, and so the possible changes are discrete. We assume that a unit enters at time t_0 into the (origin) state j, that is $Y_{t0} = j$. The basic form of change to be explained in the transition rate framework is the probability of a change in Y_t from an origin state j to a destination state k (while $t > t_0$). Now, we need a concept that allows describing the development of the process at every point in time, while the process is going on, and that, for its definition, only relies on information about the past development of the process. The crucial concept that can be used for this purpose is the <u>transition rate</u>. To define this concept, let us first introduce a random variable T to represent the duration, beginning at t_0, until a change in the dependent variable, that is a transition from (origin) state j to (destination) state k, occurs. To simplify the notation we will assume that $t_0 = 0$. Then, the following probability can be defined:

$$\Pr(t \leq t < t' \mid T \geq t) \qquad t < t' \qquad (1.2)$$

This is the probability that an event occurs in the time interval from t to t', given that no event (transition) has occurred before, that is, in the interval from 0 to t. This probability is well defined and obviously well suited to describe the temporal evolution of the process. The definition refers to each point in time while the process is evolving, and thereby can express the idea of change during its development. Also, the definition only relies on information about the past of the process, what has happened up to the present point in time, t. Therefore, the concept defined in (1) can sensibly be used to describe the process before it has finished for all individuals in the population. Assume that we know the probabilities defined in (1.2) for <u>all points in time</u> up to a certain point t^*. Then we have a description of the process up to this point, and this description is possible without knowing how the process will develop in the future, i.e. for $t > t^*$.

Since our mathematical model is based on a continuous time axis, one can in the expression (1.2) let t'- t approach zero. However, as the length of the time interval approaches zero, the concept of change in the dependent variable would simply disappear because the probability that a change takes place in an interval of zero length is zero:

$$\lim_{t' \to t} \Pr(t \leq t < t' \mid T \geq t) = 0$$

To avoid this, we regard the ratio of the transition probability to the length of the time interval to represent the probability of future changes in the dependent variable per unit of time (Coleman 1968), i.e. we consider

$$Pr(t \leq t < t' \mid T \geq t) / (t' - t)$$

This allows us to define the limit

$$r(t) = \lim_{t' \to t} Pr(t \leq t < t' \mid T \geq t) / (t' - t)$$

and we arrive at the central concept of the transition rate. This concept obviously provides the possibility of giving a local, time-related description of how the process (defined by a single episode) evolves over time. We can interpret r(t) as the propensity to change the state, from origin j to destination k, at t. But one should note that this propensity is defined in relation to a risk set, the risk set at t, i.e. the set of individuals who can experience the event because they have not already had the event before t. Having introduced the basic concept of a transition rate, we can finally formulate our basic modeling approach. The preliminary description in (1.1) can now be restated in a somewhat more precise form as

$$r(t) = g(t,x) \qquad (1.4)$$

This is the basic form of a transition rate model. The central idea is to make the transition rate, which describes a process evolving in time, dependent on time and on a set of covariates, x. Obviously, we also need the "variable" time (t) on the right-hand side of the model equation. However, it must be stressed that a sensible causal relation can only be assumed for the dependency of the transition rate on the covariates. The causal reasoning underlying the modeling approach (1.4) is

$$\Delta X_t \rightarrow \Delta r(t') \qquad t < t'$$

As a causal effect, the changes in some covariates in the past may lead to changes in the transition rate in the future, which in turn describe the propensity that the units under study will change in some presupposed state space. As discussed above, this causal interpretation requires that we take the temporal order in which the process evolves very seriously. At any given point in time, t, the transition rate r(t) can be made dependent on conditions that happened to occur in the past, i.e. before t, but not on what is the case at t or in the future after t. With respect to these individuals, and while the process is evolving, there is always a distinction in past, present, and future. This is particularly important for a causal view of the process. The past conditions the present, and what happens in the present shapes the future. There are many possibilities to specify the functional relationship g(.) in (1.4) (see Blossfeld and Rohwer 1995).

It is sometimes argued that sociologists should give up the causal analytical point of view in favor of a systems view because the operation of causal forces is mutually interdependent and variables change each other more or less simultaneously in many systems (Marini and Singer 1988). However, even in systems of interdependent processes time does not run backwards, and change in one of the interdependent variables will take (at least a small amount of) time to produce a change in another one. Thus, in systems of variables there cannot be any simultaneity of causes and their effects. This allows to demonstrate that a causal approach to interdependent systems is possible with the help of the transition rate concept (see Blossfeld and Rohwer 1995). In other words, the systems view is not a substitute for a proper causal approach in our field (Kelly and McGrath 1988).

Since the transition rate is indeed an abstraction, it is necessary to relate it back to quantities that are directly observable, that is to frequencies of state occupancies at particular points in time. To support such inferences, some additional statistical concepts are useful which are extensively described by Blossfeld and Rohwer (1995).

3. Causal Modeling and Observation Plans

Over the last 20 years, social scientists have been collecting event history data with increasing frequency. This is not an accidental trend, nor does it reflect a prevailing type of fashion in survey research. Instead, as discussed in the previous section, it indicates a growing recognition among social scientists that event history data is often the most appropriate empirical information one can get for a causal analysis.

To collect data generated by a continuous-time, discrete-state substantive process, different observation plans have been used (Coleman 1981; Tuma and Hannan 1984). With regard to the extent of detail about the process of change, one can distinguish between cross-sectional data, panel data, event count data, event sequence data, and event history data.

In this paper, we will not discuss event count data (see Barron 1993), which simply record the number of different types of events for each unit (e.g. the number of upward, downward, or lateral moves in the employment career in a period of 10 years), and event sequence data, which document the sequence of states occupied by each unit, as they are rarely used in the social sciences. On the other hand, cross-sectional data and panel data are standard sociological data types (Tuma and Hannan 1984). It is, therefore, particularly intriguing to compare event history data with cross-sectional and panel data. We will use the example shown in Figure 3. In this figure, an individual's family career is observed in a cross-sectional survey, a panel survey and an event-oriented survey.

Cross-Sectional Data

Let us first discuss the cross-sectional observation. In the social sciences, this is the most common form of data for assessing the determinants of behavior. The family history of the individual in Figure 3 is represented by one single point in time: his or her marital state at the time of interview. Thus, a cross-sectional sample is only a "snapshot" of the substantive process being studied. The point in time when researchers take that "picture" is normally not determined by hypotheses about the dynamics of the substantive process itself, but by external considerations like getting research funds, finding an appropriate institute to conduct the survey etc.

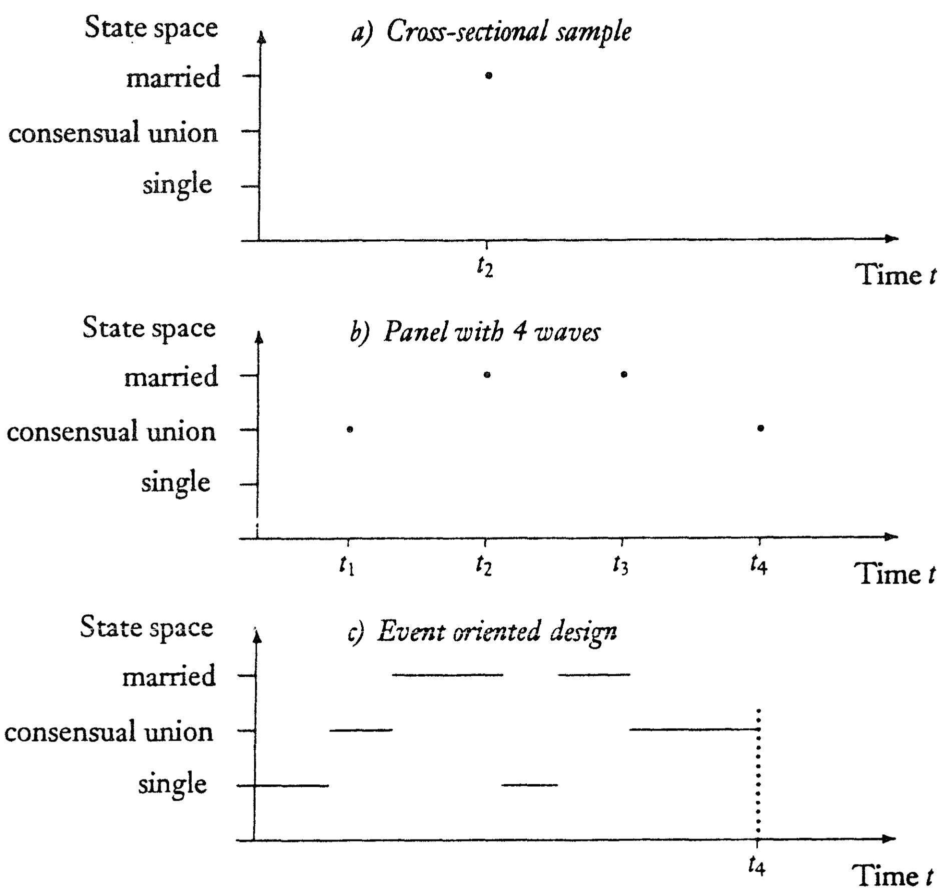

Figure 3 Observation of an individual's family career on the basis of a cross-sectional survey, a panel study and an event history oriented design.

Coleman (1981) has demonstrated that one must be cautious in drawing inferences about explanatory variables on the basis of such data because, implicitly or explicitly, social researchers must assume that the substantive process under study is in some kind of statistical equilibrium. Statistical equilibrium or stability of the process means that whilst individuals (or any other unit of analysis) may change their states over time, the state probabilities are fairly trendless or stable. Therefore, an equilibrium of the process requires that the inflows to and the outflows from each of the discrete states be equal over time to a large extent. It is only in such cases that one can reasonably assess the estimates of logit and log-linear analyses, as shown by Coleman (1981).

However, these estimates are ambiguous because they only represent the net differences in the effects of independent variables (Coleman 1981; Blossfeld and Rohwer 1995). If there is no equilibrium in the process, cross-sectional coefficients may not only be ambiguous, but even present a completely misleading picture. In a recent study on unemployment incidence, Rosenthal (1991), for example, demonstrates how confusing cross-sectional estimates can be if the unemployment rate increases or decreases in a specific region and if the process of change is, therefore, not in equilibrium.

In the social sciences one can expect that stability is very rare. For example, life history studies (Mayer 1990; Blossfeld 1989, 1995a, 1995b) show that change across age, cohort and historical period is an enduring and important feature in all domains of modern individuals' lives (Mayer and Tuma 1990); organizational studies demonstrate that most social organizations seem to follow a program of growth and not of stability; and most modern societies reveal an accelerating rate of change in almost all of their subsystems (cf. the rapid changes in family systems, job structures, educational systems etc., see Heinz 1991a, 1991b, 1992; Mayer 1990; Blossfeld 1989, 1995a, 1995b). But even in areas considered to be fairly stable, one must ask the crucial methodological question: to what extent is the process under study close to an equilibrium (Tuma and Hannan 1984)? This question can only be answered if longitudinal data are applied, as longitudinal data are the only type of data which indicate whether a steady state actually exists, or how long it will take until a system returns to a new equilibrium after some external upheaval.

Beyond the crucial assumption of process stability, cross-sectional data have many inferential limitations with regard to causal modeling. We want to address at least some of the more important problems here.[13]

Direction of causality. There are only a few situations in which the direction of causality can be established based on cross-sectional data (Davies 1987). For example, consider the strong positive association between parental socioeconomic characteristics and educational attainment of sons and daughters, controlling for other important influences (Shavit and Blossfeld 1993). A convincing interpretation of this effect might be that being born into a middle class family increases the likelihood of attaining a university degree because one is unable to think of any other plausible explanation for the statistical association. However, such recursive relationships, in which all the causal linkages run "one way" and have no "feedback" effects, are rare in social science research. For example, there is very often an association between the age of the youngest child and female labor force participation in modern industrialized societies (Blossfeld and Rohwer 1995). The common interpretation is that there is a one-way causality with young children tending to keep mothers at home. However, it is quite possible that the lack of jobs encourages women to enter into marriage and motherhood, suggesting a reversed relationship (Davies 1987).

The ambiguity of causation seems to be particularly important for the modeling of the relationship between attitudes and behavior. There are two interesting aspects of this relationship: there is a direct effect in which behavior affects attitudes, and there is a "feedback" process where attitudes change behavior (Davies 1987). Footnote: The relationship between attitudes and behavior suggests that there is some kind of inertia (or positive feedback) which means that the probability of a specific behavior increases as a monotonic function of attitudes and attitudes depend on previous behavior (Davies and Crouchley 1985). The well-known disputes among sociologists, as to whether value change engenders change in social structure, or whether structural change leads to changing values of individuals, often originates from the fact that cross-sectional surveys can only assess the net association of these two processes.

Various strengths of reciprocal effects. Connected with the inability of establishing the direction of causality in cross-sectional surveys is the drawback that these data cannot be used to discover the different strengths of reciprocal effects. For example, many demographic studies have shown that first marriage and first motherhood are closely interrelated (Blossfeld and Huinink 1991). To understand what has been happening

[13] These problems are however not mutually exclusive.

with regard to family formation in modern societies, it might be of interest not only to know the effect of marriage on birth rates, but also the effect of pregnancy or first birth on getting married (Blossfeld and Huinink 1991; Blossfeld 1995a); and perhaps, how these effects have changed over historical time (Manting 1994).

Observational data. Most sociological research is based on nonexperimental observations of social processes and these processes are highly selective. For example, Lieberson (1985) in a study examining the influence of type of school (private vs. public) on test performance among students distinguished at least three types of non-random processes: (1) there is self-selectivity, in which the units of analysis sort themselves out by choice (e.g. specific students choose specific types of schools); (2) there is selective assignment by the independent variable itself, which determines, say, what members of a population are exposed to specific levels of the independent variable (e.g. schools select their students based on their past achievement); and (3) there is also selectivity due to forces exogenous to variables under consideration at the time (socioeconomic background, ethnicity, gender, previous school career, changes of intelligence over age etc.); and many of these sources are not only not observed, but effectively unmeasurable. Although no longitudinal study will be able to overcome all the problems of identification of causal effects, cross-sectional data offer the worst of all opportunities to disentangle the effects of the causal factors of interest on the outcome from other forces operating at the same time because these data are least informative about the process of change. Cross-sectional analysis therefore requires a particularly careful justification, and the results must always be appropriately qualified (Davies 1987; Pickles and Davies 1989).

Previous history. There is one aspect of observational data that deserves special attention in the social sciences. Life courses of individuals (and other units of analysis like organizations etc.) involve complex and cumulative time-related layers of selectivity (Mayer 1991). Therefore, there is a strong likelihood that specific individuals have been entering a specific origin state. In particular, life course research has shown that the past is an indispensible factor in understanding the present (Heinz 1991a, 1991b, 1992; Mayer 1990). Cross-sectional analysis may be performed with some proxy-variables as well as with assumptions of the causal order and interdependencies between the various explanatory variables. However, it is often not possible to appropriately trace back the time-related selective processes operating in the previous history, as these data are simply not available. Thus, the normal control approaches in cross-sectional statistical techniques will rarely be successful in isolating the influence of some specific causal force (Lieberson 1985).

Age and cohort effects. Cross-sectional data cannot be used to distinguish age and cohort effects (Tuma and Hannan 1984; Davies 1987). However, in many social science applications it is of substantive importance to know whether the behavior of people (e.g. their tendency to vote for a specific party) is different because they belong to different age groups or whether they are members of different birth cohorts (Blossfeld 1986, 1989).

Historical settings. Cross-sectional data are not able to take into account the fact that processes emerge in particular historical settings. For example, in addition to individual resources (age, education, labor force experience etc.), there are at least two ways in which a changing labor market structure affects career opportunities. The first is that people start their careers in different structural contexts. It has often been assumed that these specific historic conditions at the point of entry into the labor market have a substantial impact upon people's subsequent careers. This kind of influence is generally called a cohort effect (Glenn 1977). The second way that changing labor market structure influences career opportunities is that it improves or worsens the career prospects of all people within the labor market at a given time. For example, in a favorable economic situation with low unemployment, there will be a relatively wide range of opportunities. This kind of influence is generally called a period effect (Mason and Fienberg 1985). With longitudinal data, Blossfeld (1986) has shown that lifecourse, cohort, and period effects can be identified based on substantively developed measures of these concepts (Rodgers 1982), and that these effects represent central mechanisms of career mobility that must be distinguished.

Multiple clocks, historical eras and point-in-time events. From a theoretical or conceptual point of view, multiple clocks, historical eras and point-in-time events very often influence the substantive process being studied (Mayer and Tuma 1990). For example, in demographic studies of divorce, types of clocks, such as age of respondent, time of cohabitation, duration of marriage, ages of children, as well as different phases in the state of the business cycle, or changes in national (divorce) laws are of importance (Blossfeld, DeRose, Hoem, and Rohwer 1993). With respect to cross-sectional data, such relationships can hardly be studied without making strong untestable assumptions.

Contextual processes at different levels. Social scientists are very often interested in the influences of contextual processes at different aggregation levels (Huinink 1989). Contextual process effects refer to situations where changes in the group contexts themselves influence the dependent variable. For example, career mobility of an individual may be conceptualized as being dependent on changes in resources at the individual level (e.g. social background, educational attainment, experience etc.), the success of the firm in which he/she is employed (e.g. expansion or contraction of the organization) at the intermediate level, and changes in the business cycle at the macro level (Blossfeld 1986; DiPrete 1993). Cross-sectional data do not provide an adequate opportunity for the study of such influences at different levels (Mayer and Tuma 1990).

Duration dependence. Another problem of cross-sectional data is that they are inherently ambiguous with respect to their interpretation at the level of the unit of observation. Suppose we know that in the Federal Republic of Germany, 30.6 per cent of employed women were working part-time in 1970 (Blossfeld and Rohwer 1995). At the one extreme, this might be interpreted to imply that each employed woman had a 30.6 per cent chance of being employed part-time in this year, but on the other, one could infer that 30.6 per cent of the employed women always worked part-time and 69.4 per cent were full-timers only. In other words, cross-sectional data do not convey information about the time women spent in these different employment forms. They are therefore open to quite different substantive interpretations (Heckman and Willis 1977; Flinn and Heckman 1982). In the first case, each woman would be expected to move back and forth between part-time and full-time employment. In the second, there is no mobility between part-time and full-time work, and the estimated percentages describe the proportions of two completely different groups of employed women. From an analytical point of view, it is therefore important to have data about durations in a state. Also, repeated cross-sectional analysis using comparable samples of the same population (e.g. a series of microcensuses or cross-sectional surveys), can only show net change, not the flow of individuals.

Variability in state dependencies. In many situations cross-sectional data are problematic because the rate of change is strongly state dependent and entries into and exits from these states are highly variable over time (e.g. over the life course and historical period or across cohorts). For example, it is well-known that the roles of wives and mothers (the latter in particular) have been central in women's lives. Therefore, the family cycle concept has frequently been used in sociology to describe significant changes in the circumstances which affect the availability of women for paid work outside the home. The basic idea is that there is a set of ordered stages primarily defined by variations in family composition and size that could be described with cross-sectional data. However, this view often leads to the tendency to assume that what happens to different women in various phases in the family cycle at one point in time is similar to the pattern that women experience when they make these transitions in different historical times (which has been called the "life course fallacy"). Moreover, there is the well-known problem that individuals and families often fail to conform to the assumption of a single progression through a given number of stages in a predetermined order. At least three reasons for this may exist (Murphy 1991): (1) the chronology of timing of events may not conform to the ideal model, for example childbearing may start before marriage; (2) many stages are not reached, for example, by never-married persons; and (3) the full set of stages may be truncated by events such as death or marital breakdown. Such complex constellations between the family cycle and women's labor force participation could hardly be meaningfully described or studied on the basis of cross-sectional data (see also Blossfeld 1995b).

Changes in outcomes. Cross-sectional models very often have a tendency to over-predict change and consistently over-estimate the importance of explanatory variables (Davies 1987). The reason for this phenomenon is that these analyses cannot be based on how changes in explanatory variables engender changes in outcomes. They are only concerned with how levels of explanatory variables "explain" an outcome at a specific point in time. However, if an outcome at time t (e.g. choice of mode of travel to work in June) is dependent on a previous outcome (e.g. established choice of mode of travel to work), and if both outcomes are positively influenced in the same way by an explanatory variable (e.g. merits of public transport), then the effect of the explanatory variable will reflect both the true positive influence of the explanatory variable on the outcome at time t and a positive spurious element due to that variable acting as a proxy for the omitted earlier outcome (established mode of travel to work). Thus, a cross-sectional analysis of the travel to work choice (e.g. public vs. private transport) would have a tendency to overpredict the effect of policy changes (e.g. fare increases or faster buses) because there is a strong behavioral inertia (Davies 1987).

These examples show that cross-sectional data have many severe inferential limitations for social scientists. Therefore, it is not surprising that causal conclusions based on cross-sectional data have often been radically altered after the processes were studied with longitudinal data (Lieberson 1985).

Longitudinal studies also have a much greater power than cross-sectional ones, both in the estimation of bias from missing data, and in the means for correcting it. This is because in longitudinal studies one often has data from previous points in time, thus enabling the characteristics of non-responders or lost units to be assessed with some precision. It is noteworthy that almost all the substantive knowledge concerning the biases associated with missing data, which all studies must seek to minimize, is derived from longitudinal studies (Medical Research Council 1992).

Although longitudinal data are no panacea, they are obviously more effective in causal analysis and have less inferential limitations (Magnusson, Bergmann and Torestad 1991; Arminger, Clogg and Sobel 1995; Clogg and Arminger 1993; Blossfeld 1995a, 1995b; Mayer 1990; Mayer and Tuma 1990). They are indispensable for the study of processes over the life course (of all types of units) and their relation to historical change. Therefore, research designs aimed at a causal understanding of social processes should be based on longitudinal data at the level of the units of analysis.

Panel Data

The temporal data most often available to sociologists are panel data, for which the same persons or units are re-interviewed or observed at a series of discrete points in time (Chamberlain 1984; Hsiao 1986; Arminger and Muller 1990; Engel and Reinecke 1994). Figure 3 shows four-wave panel in which the family career of the respondent was observed at four different points in time. This means that there is only information on states of the units at pre-determined survey points, but the course of the events between the survey points remains unknown.

Panel data normally contain more information than cross-sectional data, but involve well-known distortions created by the method itself (see, e.g., Magnusson and Bergmann 1990; Hunt 1985):

Panel bias. Respondents often answer many questions differently in the second and later waves than they did the first time; perhaps this is because they are less inhibited, or they mulled over or discussed the issues between questioning dates.

Modification of processes. Panels tend to influence the very phenomena they seek to observe - this changes the natural history of the processes being observed.

Regression to the mean. During the first panel wave, some panel members answer specific questions a bit differently than they normally would, due to chance circumstances at the time of measurement. For example, when a respondent has some problems during the first observation, he/she may perform poorly on some topics, but will report more accurately at the next panel wave. This phenomenon, known as "regression to the mean," can easily be mistaken by researchers for specific trends.

Attrition of the sample. In panel studies the composition of the sample normally diminishes selectively as time goes by. Therefore, what researchers observe in the panel may not give a true view of what has happened to their original sample.

Non-responses and missing data. In a cross-sectional analysis, one can afford to throw out a small number of cases with "non-responses" and "missing data," but in a long-term longitudinal study, throwing out incomplete cases at each round of observations can eventually leave a severely pruned sample having very different characteristics from the original one.

Fallacy of cohort centrism. Very often panel studies are focused on members of a specific cohort (cf., for example, the British National Child Study). In other words, these panels study respondents that were born in, grew up in, and have lived in a particular segment of history. There is therefore a danger that researchers might assume that what happens to a particular group of people over time reveals general principles of the life course (fallacy of cohort centrism). Many events may simply be specific for that generation.

Fallacy of period centrism. Many panel studies include just a few waves and, therefore, cover only a short period of historical time (cf. the German Socio-Economic Panel which now covers 10 years). At the time of a particular observation, special conditions may exist and this can result in an individual responding differently than he/she normally would (fallacy of historical period).

Confounded age, period and cohort effects. In any long-term panel study in sociology, three causal factors - individual's age, cohort, and period effect - are confounded (cf. the Panel Study of Income Dynamics). Analytical techniques are necessary to unconfound these three factors and reveal the role of each. As discussed in more detail below, panel data do have some specific problems unconfounding the three major factors. However, for gaining scientific insights into the interplay of processes governing life courses from birth to death, they appear to be a better approach than applying cross-sections. But a mechanical and atheoretical cohort analysis is a useless exercise and statistical innovations alone will not solve the age-period-cohort problem (Blossfeld 1986; Mayer and Huinink 1990).

Most of the above-mentioned difficulties concerning panel studies can be dealt with by sophisticated statistical procedures or more panel waves. However, panel data also lead to a series of deficiencies with respect to the estimation of transition rates (Tuma and Hannan 1984): first, there is the problem of "embeddability," which means that there may be difficulties in embedding a matrix of observed transition probabilities within a continuous-time Markov process (Singer and Spilerman 1976a); second, there is the problem that there may be no unique matrix of transition rates describing the data (Singer and Spilerman 1976b); and third, there is the drawback that the observed matrix of transition probabilities may be very sensitive to sampling and measurement error (Tuma and Hannan 1984). Multiple waves with irregular spacing or shorter intervals between waves can reduce these problems. However, as Hannan and Tuma

(1979) have noted, the more panel and event history data resemble each other, the less problematic modeling becomes.

Lazarsfeld (1948, 1972) was among the first sociologists to propose panel analysis of discrete variables. In particular, he wanted to find a solution to the problem of ambiguity in causation. He suggested that if one wants to know whether a variable X induces change in another variable Y, or whether Y induces change in X, observations of X and Y at two points in time would be necessary. Lazarsfeld applied this method to dichotomous variables whose time-related structure he analyzed in a resulting sixteenfold table. Later on, Goodman (1973) applied log-linear analysis to such tables. For many years, such a cross-lagged panel analysis for qualitative and quantitative variables (Campbell and Stanley 1963; Shingles 1976) was considered to be a powerful quasi-experimental method of making causal inferences. It was also extended to multiwave-multivariable panels to study more complex path models with structural-equation models (Jöreskog and Sörbom 1993). However, it appears that the strength of the panel design for causal inference was hugely exaggerated (Davis 1978). Causal inferences in panel approaches are much more complicated than has been generally realized. Several reasons are responsible for this:

Time until the effect starts to occur. As discussed above, it is important to realize that the role of time in causal explanations does not only lie in specifying a temporal order in which the effect follows the cause in time. It additionally means that a temporal interval is necessary for the cause to have an impact (Kelly and McGrath 1988). The time interval may be short or long, but can never be zero or infinity (Kelly and McGrath 1988). In most of the current sociological theories and interpretations of research findings this interval is left unspecified. In most cases, at least implicitly, researchers assume that the effect takes place almost immediately. Of course, if this is the case, then there seems to be no need for theoretical statements about the time course of causal effects. A single measurement of the effect at some point in time after a cause has been imposed might be sufficient for catching it (see Figure 1a). However, if there is a reason to assume that there is a lag between cause and effect, then a single measurement of the outcome is inadequate for describing the process (see Figure 1b); and the interpretation arrived at on the basis of that single measurement will be a function of the point in time chosen to measure the effect. Thus, a restrictive (implicit) assumption of panel designs is that either cause and effect occur almost simultaneously, or the interval between observations is of approximately the same length as the true causal lag. The greater the discrepancy, the greater the likelihood that the panel analysis will fail to discover the true causal process. Thus, as expressed by Davis (1978), if one does not know the causal lag exactly, panel analysis is not of much utility to establish causal direction or time sequencing of causal effects.

Unfortunately, we rarely, if ever, have enough information about the detailed structure of a social process to specify the true lag precisely.

Temporal shapes of the unfolding effect. While the problem of time-lags is widely recognized in sociology, there is almost no discussion with respect to the temporal shapes of effects (Kelly and McGrath 1988) (see Figure 1). If the effect increases or decreases monotonically (Figure 1c) or linearly, oscillates in cycles (Figure 1.e), or shows any other complicated time-related pattern, then the strength of the observed effect is dependent on the timing of the panel waves. A panel design might be particularly problematic if there are non-monotonic cycles of the effect, because totally opposite conclusions about the effects of the explanatory variable can be arrived at, depending on whether the panel places measurement points at a peak or at an ebb in the curve (see Figures 1d and 1e).

Reciprocal effects with different time-paths. In cases of reciprocal causality, additional problems will arise if the time structure of the effects of X_1 on X_2 and of X_2 on X_1 are different with respect to lags and shapes. In these situations, a panel design might turn out to be completely useless for those wishing to detect such time-related recursive relationships.

Observational data and timing of measurement of explanatory variables. Most sociological research is based on observational data, meaning that manipulation of the timing of the independent variables is generally not possible. For example, if the researcher is going to study the effects of job mobility on marriage behavior, it is impossible to force respondents to change their jobs, say, at the time of the first panel wave. Thus, the longer the interval between panel waves, the more uncertainty there will be regarding the exact point in time when an individual moved to another job and therefore about the point we evaluate in the time-path of the effect (Coleman 1981). The situation may be even more problematic if changes in independent variables are repeatable and several changes are possible between two successive panel waves, as might be the case with job shifts observed in yearly intervals (cf. Sandefur and Tuma 1987). In such panel studies, even the causal order of explanatory and dependent events may become ambiguous.

Observational data and the timing of control variables. Observational studies take place in natural settings and, therefore, offer little control over changes in other important variables and their timing. If these influences change arbitrarily and have time-related effect patterns, then panel studies are useless in disentangling the effect of interest from time-dependent effects of other parallel exogenous processes.

Continuous changes of explanatory and control variables. In observational studies, explanatory and control variables may not only change stepwise from one state to another, but can often change continuously over time. For example, individuals continuously change age, constantly acquire general labor force experience or job-specific human capital if they are employed (Blossfeld and Huinink 1991), are exposed to continuously changing historical conditions (Blossfeld 1986), are steadily changing their social relationships in marital or consensual unions (Blossfeld, De Rose, Hoem, and Rohwer 1993), etc. Even in cases where these continuous changes are not connected with lags or time-related effect patterns, there are deficiencies of panel data concerning their capabilities of detecting time dependence in substantive processes. This is why panel analysis can often not appropriately identify age, period, and cohort effects (Blossfeld 1986).

Therefore, the use of panel data causes an identification problem due to omitted factors whose effects are summarized in a disturbance term. These factors are not stable over time, which means that the disturbance term cannot be uncorrelated with the explanatory variables. Panel analysis therefore critically depends on solutions to the problem of autocorrelation. This problem can be reasonably well tackled by increasing the number of panel waves and modifying their spacing. Panel analysis is particularly sensitive to the length of the time intervals between waves relative to the speed of the process (Coleman 1981). They can be too short, so that too few events will be observed, or too long, so that it is difficult to establish a time-order between events (Sandefur and Tuma 1987). A major advantage of the continuous-time observation design in event history analysis is therefore that it makes the timing between waves irrelevant (Coleman 1968).

Event History Data

For many processes in the social sciences, a continuous measurement of qualitative variables seems to be the only adequate method of assessing empirical change. This is achieved by utilizing an event oriented observation design which records all the changes in qualitative variables and their timing. As shown in Figure 3, the major advantage of event history data is that they provide the most complete data possible on changes in qualitative variables which may occur at any point in time. The observation of events therefore provides an attractive alternative to the observation of states for social scientists.

Event history data, mostly collected retrospectively via life history studies, covers the whole life course of individuals. An example for such a study is the German Life

History Study (GLHS, Mayer and Brückner 1989). Retrospective studies have the advantage of normally being cheaper to collect than panel data. They are also systematically coded to one framework of codes and meanings (Dex 1991). But retrospective (in contrast to prospective) studies suffer from several limitations that have been increasingly acknowledged (Medical Research Council 1992):

Non-factual data. It is well-known that retrospective questions concerning motivational, attitudinal, cognitive, or affective states are particularly problematic because the respondents can hardly recall the timing of changes in these states accurately (Hannan and Tuma 1979). This type of data is not verifiable even in principle, as these states exist only in the minds of the respondents and are only directly accessible, if at all, to the respondent concerned (Sudman and Bradburn 1986). For these non-factual data, panel studies have the advantage of being able to repeatedly record current states of the same individual over time. Thus, for studies aiming to model the relationship between attitudes and behavior over time, panel observations of attitudinal states, combined with retrospective information on behavioral events since the last sweep, appear to be an appropriate design.

Recall problems with regard to behavior or facts. Behavioral or factual questions ask the respondents about characteristics, things they have done, or things that have happened to them which in principle are verifiable by an external observer. Most surveys (cross-sectional, panel or event oriented) elicit retrospective information on behavior or facts (e.g. by asking people about their education, social origin etc.), so that the disadvantages of retrospection are only a matter of degree. However, event history studies are particularly ambitious (see Mayer and Brückner 1989). They try to collect continuous records of qualitative variables that have a high potential for bias because of their strong reliance on memory. However, research on the accuracy of retrospective data shows that individuals' marital and fertility histories, family characteristics and education, health service usage, and employment history can be collected to a reasonable degree of accuracy. A very good overview concerning the kinds of data which can be retrospectively collected, the factors affecting recall accuracy, and the methods improving recall has been presented by Shirley Dex (1991).

Unknown factors. Retrospective designs cannot be used to study factors involving variables that are not known to the respondent (e.g. emotional and behavioral problems when the respondent was a child). In such cases, panel studies are indispensable (Medical Research Council 1992).

Limited capacity. There is a limit to respondents' tolerance for the amount of data which can be collected on one occasion (Medical Research Council 1992). A carefully corsetted panel design can therefore provide a broader coverage of variables (if these are not unduly influenced by variations at the time of assessment).

Only survivors. Due to their nature, retrospective studies must be based on survivors. Thus, those subjects who have died or migrated from the geographical area under study will necessarily be omitted. If either is related to the process (as often may be the case), biases will arise. This problem is particularly important for retrospective studies involving a broad range of birth cohorts, like the GLHS or international migration studies (Blossfeld 1987).

Misrepresentation of specific populations. Retrospective studies also systematically misrepresent specific populations. For example, Duncan (1966) has shown that if men are asked about their fathers, men from earlier generations who had no sons, or whose sons died or emigrated are not represented in a retrospective father-son mobility table.

4. Conclusion

The aim of this paper has been to show that - compared to alternative observation plans - the collection of event history data is an extremely useful approach for uncovering causal relationships or mapping out systems of causal relations. Event history models provide a natural basis for a causal understanding of social processes because they relate the change in future outcomes to conditions in the past at each point in time and enable the researcher to predict future changes on the basis of past observations at each moment of the process (Aalen 1987).

In the past, event history data normally have been collected retrospectively leading to specific problems described above. To avoid these problems (or to diminsh their relevance), a mixed design employing a follow-up (or "catch-up") and a follow-back strategy appears to combine the strengths of panel designs with the virtues of retrospective studies. Therefore, in modern panel studies event histories are collected retrospectively for the period before the panel started and between the successive panel waves. Sometimes, complete administrative records also contain time-related information about events in the past. In any case, with regard to cross-sectional and tranditional panel data, all of these procedures (retrospective, combined follow-up and back-up, or available registers) offer a comparatively superior opportunity for modeling social processes, regardless of which method is selected.

References

Aalen, O. O. (1987): "Dynamic modelling and causality," in: Scand. Actuarial Journal, 12:177-190.

Arminger, G., C.C. Clogg, and M.E. Sobel (eds.) (1995): Handbook of statistical modeling for the social and behavioral sciences, Plenum.

Arminger, G. and F. Müller (1990): Lineare Modelle zur Analyse von Paneldaten, Opladen.

Barron, D. N. (1993): "The analysis of count data: overdispersion and autocorrelation," in: Sociological Methodology, pp. 179-220.

Blossfeld, H.-P. (1986): "Career opportunities in the Federal Republic of Germany: a dynamic approach to the study of life-course, cohort, and period effects," in: European Sociological Review, 2:208-225.

Blossfeld, H.-P. (1987): "Labor market entry and the sexual segregation of careers in the FRG," in: American Journal of Sociology, 93:83-118.

Blossfeld, H.-P. (1989): Kohortendifferenzierung und Karriereprozeß. Eine Längsschnittstudie über die Veränderung der Bildungs- und Brerufschancen im Lebenslauf, Frankfurt a. M. and New York.

Blossfeld, H.-P. (ed.) (1995a): Family formation in modern societies and the new role of women, Boulder et al. (in print).

Blossfeld, H.-P. (ed.) (1995b): Between equalization and marginalization. Part-time working women in Europe and the United States of America, Oxford (forthcoming).

Blossfeld, H.-P., A. DeRose, J. Hoem, and G. Rohwer (1993): Education, modernization, and the risk of marriage disruption: differences in the effect of women's educational attainment in Sweden, West-Germany, and Italy," Stockholm Research Reports in Demography, no. 76, Stockholm (To appear in: K. Oppenheim Mason and An-Magritt Jensen (eds.): Gender and family change in industrialized countries, Oxford).

Blossfeld, H.-P., A. Hamerle and K. U. Mayer (1989): Event history analysis, Hillsdale (N.J.).

Blossfeld, H.-P. and J. Huinink (1991): "Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation," in: American Journal of Sociology, 97:143-168.

Blossfeld, H.-P., D. Manting, and G. Rohwer (1993): "Patterns of change in family formation in the Federal Republic of Germany and the Netherlands: some consequences for solidarity between generations", in: H. A. Becker and P. L. J. Hermkens (eds.): Solidarity of generations. Demographic, economic and social change, and its consequences, Amsterdam. pp. 175-196.

Blossfeld, H.-P. and G. Rohwer (1994): "West-Germany," in: H.-P. Blossfeld (ed.): Between equalization and marginalization. Part-time working women in Europe and the United States of America, Oxford (forthcoming).

Blossfeld, H.-P. and G. Rohwer (1995): Techniques of event history modeling. New approaches to causal analysis, Hillsdale (NJ) (in print).

Bollen, K. A. (1989): Structural equations with latent variables, New York.

Campbell, D. T. and J. C. Stanley (1963): Experimental and quasi-experimental designs for research, Chicago.

Campbell, R. T., E. Mutran, and R. Nash Parker (1987): "Longitudinal design and longitudinal analysis. A comparison of three approaches," in: Research on Aging, 8:480-504.

Chamberlain, G. (1984): "Panel data," in. Z. Griliches and M. D. Intriligator (eds.): Handbook of econometrics, vol. II, pp. 1247-1317. Amsterdam.

Clogg, C. C. and G. Arminger (1993): "On strategy for methodological analysis", in: Sociological Methodology, 23:57-74.

Coleman, J. S. (1968): "The mathematical study of change," in: Hubert, M. and A. B. Blalock (eds.): Methodology in Social Research, New York, pp. 428-478.

Coleman, J. S. (1981): Longitudinal data analysis, New York.

Coleman, J. S. (1990): Foundations of social theory, Cambridge.

Coleman, J.S. and L. Hao (1989): "Linear systems analysis: macrolevel analysis with microlevel data," in: Sociological Methodology, 19:395-422.

Davies, R. B. (1987): "The limitations of cross-sectional analysis," in: Crouchley, R. (ed.): Longitudinal data analysis, Aldershot et al. pp. 1-15.

Davies, R. B. and R. Crouchley (1985): "Control for omitted variables in the analysis of panel and other longitudinal data," in: Geographical Analysis, 17:1-15.

Davis, J. A. (1978): "Studying categorical data over time," in: Social Science Research, 7:151-179.

Dex, S. (ed.)(1991): Life and work history analyses: Qualitative and quantitative developments, London and New York.

DiPrete, T.A. (1993): "Industrial restructuring and the mobility response," in: American Sociological Review, 58:74-96

Duncan, O.D. (1966): "Methodological issues in the analysis of social mobility," in: N. Smelser and S.M. Lipset (eds.): Social structure and social mobility, Chicago.

Eells, E. (1991): Probabilistic causality, Cambridge.

Engel, U. and J. Reinecke (1994): Panelanalyse, Berlin.

Faulbaum, F. and P. M. Bentler (1994): "Causal modeling: some trends and perspectives," in. I. Borg and P.O. Mohler (eds.): Trends and perspectives in empirical social research, pp. 224-249, Berlin.

Flinn, Ch. J. and J. J. Heckman (1982): "Models for the analysis of labor force dynamics," in: R. Basmann and G. Rhodes (eds.): Advances in econometrics, Greenwich (C.T.), pp. 35-95.

Freeman, R. A. (1991): "Statistical analysis and shoe leather," in: Sociological Methodology, 21:291-313.

Glenn, N.D. (1977): Cohort analysis, Beverly Hills (C.A.).

Goldthorpe, J. H. (1991): "The use of history in sociology: reflections on some recent tendencies," in: Brithish Journal of Sociology, 42:211-230.

Goldthorpe, J. H. (1994): The quantitative analysis of large-scale data-sets and rational action theory: For a sociological alliance," in: Oxford, Nuffield College (mimeo).

Goodman, L.A. (1973): "Causal analysis of data from panel studies and other kind of surveys," in: American Journal of Sociology, 78:1135-1191.

Hannan, M. T. and N.B. Tuma (1979): "Methods for temporal analysis," in: Ann. Rev. Sociol., 5:303-328.

Heckman, J.J. and R.J. Willis (1977): "A beta-logistic model for the analysis of sequential labor force participation by married women," in: Journal of Political Economy, 85:27-58.

Heinz, W. R. (ed.) (1991a): Theoretical advances in life course research. Status passages and the life course, vol. I, Weinheim.

Heinz, W. R. (ed.) (1991b): The life course and social change. Status passages and the life course, vol. II, Weinheim.

Heinz, W. R. (ed.) (1992): Institutions and gate keeping in the life course. Status passages and the life course, vol. III, Weinheim.

Holland, P. W. (1986): "Statistics and causal inference," in: Journal of the American Statistical Association, 81:945-960.

Hsiao, C. (1986): Analysis of panel data. Cambridge.

Huinink, J. (1989): Mehrebenenanalyse in den Sozialwissenschaften, Wiesbaden.

Hunt, M. (1985): Profiles of social research. The scientific study of human interactions, New York.

Jöreskog, K. G. and D. Sörbom (1993): LISREL 8. Structural equation modeling with the SIMPLIS command language, Chicago.

Kelly, J. R. and J. E. McGrath (1988): On time and method, Newbury Park et al.
Lazarsfeld, P.F. (1948): "The use of panels in social research," in: Proceedings of the American Philosophy of Sociology, 92:405-410.

Lazarsfeld, P.F. (1972): "Mutual relations over time of two attributes: a review and integration of various approaches," in: M. Hammer, K. Salzinger, and S. Sutton (eds.): Psychopathology, New York.

Lieberson, S. (1985): Making it count. The improvement of social research and theory. Berkeley et al.

Lieberson, S. (1991): "Small n's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases," in: Social Forces, 70:307-320.

Magnusson, D. and L. R. Bergmann (eds.) (1990): Data quality in longitudinal research, Cambridge.

Magnusson, D., L. R. Bergmann, and B. Törestad (eds.) (1991): Problems and methods in longitudinal research: stability and change, Cambridge.

Manting, D. (1994): Dynamics in marriage and cohabitation. An inter-temporal, life course analysis of first union formation and dissolution, Amsterdam.

Marini, M. M. and B. Singer (1988): "Causality in the social sciences," in: C. C. Clogg (ed.): Sociological Methodology, pp. 347-409.

Mason, W.M. and S.E. Fienberg (1985) (eds.): Cohort analysis in social research.

Mayer, K.U. (ed.)(1990): Lebensverläufe und sozialer Wandel, Sonderheft 31, Kölner Zeitschrift für Soziologie und Sozialpsychologie, Opladen.

Mayer, K. U. (1991): "Life courses in the welfare state," in: W. R. Heinz (ed.): Theoretical advances in life course research, vol. I. pp. 171-186, Weinheim.

Mayer, K. U. and E. Brückner (1989): Lebensverläufe und Wohlfahrtsentwicklung. Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1929-31, 1939-41, 1949-51, Materialien aus der Bildungsforschung, no. 35, Max-Planck-Institut für Bildungsforschung, Berlin.

Mayer, K. U. and J. Huinink (1990): "Age, period, and cohort in the study of the life course: a comparison of classical A-P-C-analysis with event history analysis or farewell to Lexis?," in. D. Magnusason and L. R. Bergmann (eds.): Data quality in longitudinal research, pp. 211-232, Cambridge.

Mayer, K. U. and N. B. Tuma (eds.) (1990): Event history analysis in life course research, Madison (W.I.)

Medical Research Council (1992): Review of longitudinal studies, London (mimeo).

Murphy, M. (1991): The family life cycle, London School of Economics, London (mimeo).

Pickles, A. R. and R. B. Davies (1989): "Inference from cross-sectional and longitudinal data for dynamic behavioural processes," in: J. Hauer et al. (eds.): Urban dynamics and spatial choice behaviour, pp. 81-104.

Rodgers, W.L. (1982): "Estimable functions of age, period, and cohort effects,", in: American Sociological Review, 47:774-787.

Rosenthal, Leslie (1991): "Unemployment incidence following redundancy; the value of longitudinal approaches," in: Dex, S. (ed.): Life and work history analyses: Qualitative and quantitative developments, London and New York, pp. 187-213

Sandefur, G. D. and N. B. Tuma (1987): "How data type affects conclusions about individual mobility," in: Social Science Research, 16:301-328.

Shavit, Y. and H.-P. Blossfeld (eds.): Persistent inequality. Changing educational attainment in thirteen countries, Boulder et al.

Shingles, R. D. (1976): "Causal inference in cross-lagged panel analysis," in: Political Methodology, 3:95-133.

Singer, B. and S. Spilerman (1976a): "The representation of social processes by Markov models," in: American Journal of Sociology, 82:1-54.

Singer, B. and S. Spilerman (1976b): "Some methodological issues in the analysis of longitudinal surveys," in: Annals of Economic and Social Measurement, 5:447-474.

Sudman, S. and N. M. Bradburn (1986): Asking questions, San Francisco et al.

Tuma, N. B. and M. T. Hannan (1984): Social dynamics. Models and Methods. Orlando et al.

Yamaguchi, K. (1991): Event history analysis, Newbury Park (C.A.) et al.